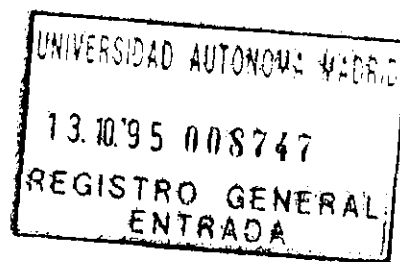


12813 230

Universidad Autónoma de Madrid
Facultad de Ciencias Económicas y Empresariales
Dpto. de Financiación e Investigación Comercial

EL ALCANCE Y LA DISTRIBUCIÓN DE CONTACTOS PUBLICITARIOS



X 541626816

R.º. FEE. 61700
M

TESIS DOCTORAL

Presentada por: Ignacio Redondo Bellón
Dirigida por: Prof. Dr. D. Ignacio Cruz Roche
Madrid, noviembre de 1995

A MIS PADRES

.

Agradezco la dirección de esta Tesis a su Director **D. Ignacio Cruz Roche** por sus orientaciones y claros consejos; por su insistencia en que este trabajo tuviera aplicaciones prácticas; y por su talante abierto a sugerencias de otros profesores, especialmente las facilitadas por **D^a. M^a Ángeles González Lobo**, quien me inició en la *planificación de medios* durante la Licenciatura; que me sugirió la investigación de la fórmula de Agostini; y que me presentó a **D. Enrique Carrero** (ex-Presidente de AIMC-EGM) y a **D. Luis Pistoni Durán** (Director Comercial de ODEC), quienes me permitieron la utilización gratuita de las fuentes de datos y de los programas de planificación, respectivamente. Debo, a **D. Julián Bravo Navalpotro** (Presidente Ejecutivo de AIMC-EGM), el presentarme a prestigiosos profesionales y el constante apoyo. A **D. Carlos Lamas** (AIMC-EGM), el abrirme nuevos horizontes en mis investigaciones. A **D^a Ana Gallardo** (ODEC), el ayudarme a manejar los programas TOM. Y a **D. Luis Ángel Sanz de la Tajada** (Universidad Complutense de Madrid), **D. Tomás Sereno** (ODEC), **D. Kevin Holland** (HolosMedia) y **D. George Lawrence** (Media Hot Line), su acogida y sus consejos. Sin todos los arriba citados, esta tesis no hubiera llegado a término. Mi más sincero agradecimiento a todos ellos.

ÍNDICE

INTRODUCCIÓN	11
I PARTE: PRÁCTICA ACTUAL DE LA PLANIFICACIÓN	23
A) INVERSIÓN PUBLICITARIA	25
B) ESTUDIO GENERAL DE MEDIOS	29
C) PROGRAMAS DE PLANIFICACIÓN TOM	36
1. Programas disponibles en Mainframe y en Tom-Micro	39
TOM0 Cuantificación de públicos objetivos	39
TOM1 Clasificación de soportes	40
TOM1D Duplicación entre soportes	41
TOM2 Evaluación de planes	43
Proceso de estimación del alcance y la distribución de contactos	44
TOM2C/CX Evaluación de planes con distribución temporal	50
TOM2M Evaluación multimedia	52
TOM3 Optimización de planes	53
TOM4 Generación de planes óptimos	56
TOM5 Consultas a fichero: tabulaciones	56
2. Programas disponibles únicamente en Mainframe	57
TOMCOB Análisis de Coberturas Inserciones	57
TOMSEG Segmentación	58
Otros programas TOM	59
D) CONTRASTE FRENTE A UN PANEL	61
1. Descripción del experimento e hipótesis implícitas	62
2. Precisión del alcance estimado	71
3. Precisión en la estimación de la distribución de contactos	74
II PARTE: LA AUDIENCIA NETA	83
INTRODUCCIÓN	85
A) MODELOS AD-HOC	87
Conceptos preliminares	87
1. La fórmula de Agostini	88
1.1. Planteamiento del autor	88
1.2. Ejemplo	92
1.3. Investigaciones posteriores	94
1.3.1. Aplicaciones en Estados Unidos y Canadá	94
1.3.2. Aplicaciones en el Reino Unido	95
1.3.3. Aplicaciones en Francia	96
1.3.4. Aportación de Claycamp y McClelland	98
1.4. Investigación realizada en nuestro país	103
1.4. Resultados de la investigación	110
1.4.1. Aplicación de la fórmula de Agostini	110
1.4.2. Análisis a posteriori del modelo de Agostini	116
1.4.3. Aplicación con el parámetro k variable	123

2. La fórmula de Kuhn	130
2.1. Desarrollo del autor	130
2.2. Ejemplo	132
2.3 Aplicación a nuestro estudio	133
3. La fórmula de Kaatz	135
3.1. Planteamiento original y adaptaciones	135
3.2. Ejemplo	138
3.3. Test de la fórmula modificada	139
4. Modelo de Hofmans	142
4.1. Fundamentos e hipótesis	142
4.2. Ejemplo	146
4.3. Aplicación a la investigación propuesta	148
5. Modelo de Starch	151
5.1. Hipótesis y desarrollo	151
5.2. Ejemplo	152
5.3. Tests sobre las campañas propuestas	153
6. Estimación de duplicaciones por Little y Lodish	156
6.1. Fundamentos e hipótesis	156
6.2. Test de las duplicaciones estimadas	157
6.3. Fórmulas alternativas	161
6.3.1. Ecuación geométrica	162
6.3.2. Ecuación logarítmica sin término independiente	163
6.3.3. Ecuación logarítmica con término independiente	163
6.3.4. Ecuación exponencial	164
6.3.5. Ecuación lineal sin término independiente	164
6.3.6. Ecuación lineal con término independiente	165
6.4. Conclusiones	166
7. Modelo de Lozano	167
B) MODELOS ESTOCÁSTICOS	168
Conceptos previos e hipótesis fundamentales	168
1. Modelo binomial	171
1.1. Definición de hipótesis y funcionamiento	171
1.2. Ejemplo	173
1.3. Tests del alcance neto estimado	174
1.4. Contrastes de la distribución de contactos	176
2. Modelo de Sainsbury	179
2.1. Hipótesis y descripción	179
2.2. Ejemplo	181
2.3. Aplicación al alcance neto de nuestras campañas	182
3. Modelo de Sainsbury modificado	184
3.1. Modificación y nuevo desarrollo	184
3.2. Ejemplo	186
3.3. Aplicación al alcance neto de nuestras campañas	187
4. Modelo de Kwerel	190
4.1. Hipótesis y razonamientos	190
4.2. Ejemplo	191
4.3. Tests del alcance neto	193

5. Modelos de duplicaciones <i>entre-soportes</i>	195
5.1. Modelo de Goodhardt y Ehrenberg	195
Ejemplo	197
5.2. Modelo de Headen-Klomp-maker-Rust	199
5.3. Modelo de Rust-Zimmer-Leone	201
6. Modelos de agregación secuencial	204
6.1. Desarrollo e hipótesis implícitas	204
6.2. Ejemplo	207
7. Modelo de independencia condicional	210
7.1. Hipótesis y estructura interna	210
7.2. Ejemplo	212
8. Modelo de Hofmans ampliado	214
8.1. Extensión a la distribución de contactos	214
8.2. Ejemplo	216
9. Modelo de Metherringham	218
9.1. Fundamentos teóricos y desarrollo	218
9.2. Ejemplo	224
9.3. Tests del alcance neto estimado	226
9.4. Contrastes sobre la distribución de contactos	229
C) ANÁLISIS COMPARATIVO Y CONCLUSIONES	232
III PARTE: LA AUDIENCIA ACUMULADA	237
INTRODUCCIÓN	239
A) MODELOS AD-HOC	242
1. Modelo Logarítmico	243
2. Modelo Exponencial	245
3. Modelo de Agostini	247
4. Modelo de Engelman	251
5. Modelo de Hofmans	254
B) MODELOS ESTOCÁSTICOS	259
1. Modelo Binomial	261
2. Modelo de heterogeneidad lineal	263
Ejemplo	266
3. Modelo Beta Binomial	268
Ejemplo	271
C) MODELOS DE SIMULACIÓN	273
1. Método actuarial	274
2. Método de la probabilidad fija	276
3. Método empírico	277
4. Método de la probabilidad personal	278
5. Método del último periodo	280

PARTE IV: LA AUDIENCIA NETA ACUMULADA	283
INTRODUCCIÓN	285
A) MODELOS ESTOCÁSTICOS	286
Conceptos previos e hipótesis implícitas	286
1. Modelo Binomial	288
2. Modelo Binomial Multivariable con soportes independientes	290
2.1. Estructura y desarrollo	290
2.2. Ejemplo	292
2.3. Aplicación a nuestra investigación	293
3. Modelo Multinomial	296
3.1. Fundamentos e hipótesis	296
3.2. Ejemplo	297
4. Modelo de Markov-Bernoulli	300
4.1. Hipótesis y desarrollo	300
4.2. Ejemplo	303
5. Modelo de Metherringham	306
5.1. Hipótesis y estructura	306
5.2. Ejemplo	310
6. Modelo Matriz Beta	312
7. Modelo Multinomial Compuesto	316
7.1. Estructura básica	316
7.2. Ejemplo	320
8. Modelo de Hofmans	324
8.1. Explicación integral del fenómeno	324
8.2. Ejemplo	327
9. Modelo de Morgensztern	330
9.1. Descripción del proceso	330
9.2. Ejemplo	332
10. Modelo de Casualidad Constante	334
10.1. Fundamentos teóricos	334
10.2. Ejemplo	336
B) MODELOS DE SIMULACIÓN	338
Conceptos previos y nueva perspectiva	338
1. Modelo CAM	341
2. Modelo SCAL	346
3. Modelo Ad-Me-Sim	350
4. Modelo de Friedman	354
5. Modelo ADMOD	359
6. Modelo Calculus-Mediac	365
CONCLUSIONES	375
BIBLIOGRAFÍA	381
Advertencias sobre el diskette anejo	393

INTRODUCCIÓN

.

La **planificación de medios** es la actividad publicitaria dirigida a seleccionar los medios de comunicación y los soportes específicos, para transmitir un mensaje a un determinado público objetivo, en función de los objetivos comerciales y de los propios de la campaña. Se pretende obtener el mejor rendimiento de un presupuesto generalmente escaso, buscando la combinación de medios/ soportes que consiga mayor eficacia al menor coste posible. Es una parte del proceso publicitario, y muchas decisiones suelen venir impuestas por fases anteriores: el gasto final en medios, la duración de la campaña, la forma creativa, etc.

La selección del medio se lleva a cabo atendiendo a diversos criterios¹: unos son cualitativos, como las características del medio (imagen, color, sonido, movimiento,...) y la adecuación del medio al producto; otros cuantitativos, como el montante presupuestario (p.ej., la TV requiere una inversión mínima), el alcance del medio (mayoritario, específico,...), el precio de la inserción y la afinidad (entre la audiencia y el público objetivo); y, por último, algunos tienen parte cualitativa y parte cuantitativa, como las restricciones en otros medios (p.ej., legales), la utilización por la competencia y la estacionalidad. La selección del soporte se puede basar en criterios de cobertura, rentabilidad o afinidad, en factores cualitativos, etc.

Según Sissors y Bumba², este proceso exige previamente [1] el análisis de la situación de mercado (tamaño y cuota de mercado, análisis de los costes-beneficios, formas de distribución,...), [2] la definición del consecuente plan estratégico de marketing (objetivos de marketing, estrategias de producto, precio,...) y [3] la concreción del plan estratégico creativo (posicionamiento publicitario, objetivos específicos, contenido,...). Después el planificador debe fijar los *objetivos de medios*, determinar el/los segmento/s posible/s comprador/es, definir dónde y cuándo concentrar

¹ GONZÁLEZ LOBO, M^a Ángeles. "Curso de Publicidad". Eresma & Celeste Ediciones. Madrid. 1994. Págs. 319-23.

² SISSORS, Jack Z. and BUMBA, Lincoln: "Advertising Media Planning". NTC Business Books. NTC Publishing. 3^a Edition. 1990. Págs. 6-10.

la publicidad. La *estrategia de medios* es el conjunto de decisiones sobre alternativas diversas, para alcanzar los objetivos de medios (cfr. **Figura 1.1**).

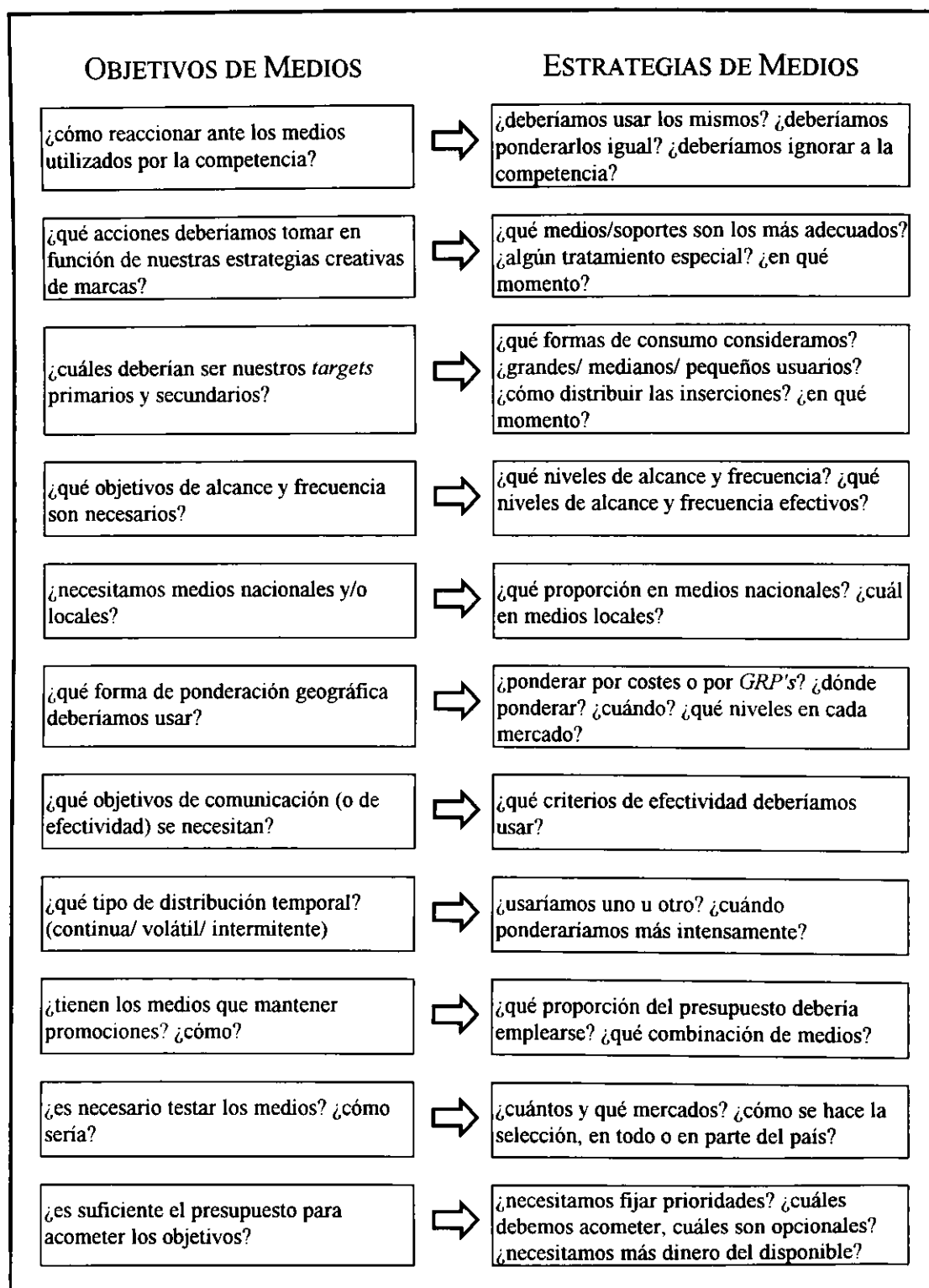


Figura 1.1: Tipos de preguntas para tomar decisiones acerca de los objetivos y estrategias de medios. Fuente: Sissors y Bumba (op.cit.).

La evaluación de las combinaciones de medios/soportes se puede llevar a cabo con múltiples factores. Tales se investigaron en dos encuestas postales³, separadas por una década; se preguntaba a una muestra de directores de medios de las 200 mayores agencias de Estados Unidos: en la primera se recibieron 91 cuestionarios válidos, y en la segunda 63. La respuesta era semicerrada, con 11 factores dados, mas los personalmente añadidos en "Otros". Se les pedía valorar cada factor en una escala, desde 1 como "nada importante" hasta 5 como "muy importante"; en la **Tabla 1.1**, se recoge el porcentaje de utilización y la valoración media según la anterior escala.

FACTORES	Utilización		Valoración	
	1982	1993	1982	1993
Alcance	93,4%	81,0%	4,4	4,5
Coste Por Mil	89,0%	77,8%	4,5	4,2
Factores cualitativos	87,9%	77,8%	4,4	4,1
Coste total	85,7%	77,8%	4,3	4,3
Distribución de Frecuencias	84,6%	74,6%	4,0	4,2
GRP's	91,2%	73,0%	4,2	4,1
Frecuencia media	90,1%	73,0%	4,2	4,0
Alcance efectivo	59,3%	68,3%	4,4	4,5
Continuidad	70,3%	65,1%	3,8	3,8
Quintiles	63,7%	58,7%	3,0	3,3
Distrib. inversión por medios	53,8%	50,8%	3,8	3,4
Otros	9,9%	11,1%	4,0	4,4

Tabla 1.1: Usos y ponderaciones de los factores de evaluación del plan.

Fuente: Leckenby y Kim (op.cit.), elaboración propia.

Veamos a continuación cómo se definen tales factores:

- Alcance (*reach*): conjunto de individuos del público objetivo (*target*) que reciben al menos un impacto del plan de medios (*media plan*); cuando se mide en porcentaje sobre el público objetivo, se denomina cobertura (*coverage*). Con los

³ LECKENBY, John D. and BOYD, Marsha M. "How Media Directors View Reach/Frequency Model Evaluation Standards". *Journal of Advertising Research*. Vol. 24. Nº 5. Octubre/Noviembre 1984. Págs. 43-52.

LECKENBY, John D. y KIM, Heejin. "How Media Directors View Reach/Frequency Estimation: Now and Decade Ago". *Journal of Advertising Research*. Vol. 34. Nº 5. Septiembre-Octubre de 1994. Págs. 9-21.

términos alcance bruto (*gross reach*) se denomina el número total de impactos del plan de medios. El impacto, exposición o contacto, se da cuando una persona del público objetivo se expone a un soporte o vehículo (*vehicle*) que incluye una inserción de la campaña; puede suceder que individuo se exponga al soporte pero no al anuncio (leer una publicación no implica ver todos los anuncios; tener encendida la TV no significa necesariamente fijar la atención durante los anuncios; etc). Entonces sería más apropiado hablar de "oportunidades de ver" (*opportunities to see, OTS*) u "oportunidades de oír" (*opportunities to hear, OTH*).

- *Gross Rating Points, GRP's*: cociente entre el número de impactos y el público objetivo; dicho porcentaje se asimila a la cobertura bruta, pero esos términos no se usan habitualmente; también es el producto de la cobertura y la frecuencia media.

- Coste por mil (*CPM*): índice de la rentabilidad que mide el gasto de 1000 impactos ó 1000 impactados; resulta del cociente entre el presupuesto y los impactos/impactados, en %.

- Distribución de contactos o de frecuencias (*frequency distribution*): distribuye la población objetivo según el número de contactos recibidos (0, 1, 2,..., hasta las n inserciones de la campaña); habitualmente las frecuencias se expresan en porcentaje.

- Frecuencia media (*average frequency*): es la media de la distribución de frecuencias; resulta de dividir los impactos entre los impactados; es el número de exposiciones recibidas por término medio en la población.

- Alcance efectivo (*effective reach*): parte del alcance que recibe los impactos adecuados para asegurar la posible respuesta (suficientes para memorizar el mensaje y no demasiados para no producir hastío). Todas las magnitudes anteriores se pueden expresar en términos de efectividad. Los umbrales inferior y superior se definen de

distinta forma según el producto, el medio, el público objetivo, el anuncio, etc.; incluso los planificadores no definen igual el alcance efectivo⁴.

- Distribución por quintiles (*quintile distribution*): es un concepto afín a la distribución de contactos, pero la población objetivo se distribuye en cinco partes iguales, a las que se asocia la frecuencia media correspondiente; aunque menos completa que la distribución de contactos, facilita la evaluación y permite comparar distribuciones de distintos tamaños.

La publicidad, en su sentido actual, nace cuando comienzan los medios de comunicación, y toma importancia cuando se produce en masa y se quiere generar un consumo en masa; en sentido lato se pueden encontrar precedentes antiquísimos de la publicidad. Del mismo modo, la planificación de medios se origina cuando se dispone de fuentes de información de esos medios y llega a su apogeo cuando se dispone de computadoras. Igual que una decisión no puede tomarse sin información previa, hoy en día la planificación de medios debe basarse en el análisis sistemático de las fuentes de datos, mediante los instrumentos de cálculo disponibles.

El Alcance y la Distribución de Contactos Publicitarios no son un fin en sí mismo, sino una etapa en la determinación de las curvas de respuesta, y una medida práctica para la evaluación, comparación y, es su caso, optimización de campañas⁵. Históricamente la literatura ha distinguido estos dos enfoques: [1] el de *evaluación*, para medir los factores que caracterizan una combinación de medios y [2] el de *optimización*, para elegir la mejor combinación de medios, habitualmente a partir de las medidas aportadas por la evaluación. Este trabajo se limita a la evaluación, aunque haya alguna referencia a la optimización. Varias razones nos movieron al presente estudio:

⁴ KRESHEL, Peggy J.; LANCASTER, Kent M.; and TOOMEY, Margaret A. "How Leading Advertising Agencies Perceive Effective Reach and Frequency". *Journal of Advertising*. Vol. 14. Nº 3. 1985. Págs. 32-8.

⁵ MARC, M.: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Pág. 11.

a) El **alcance** y la **distribución de contactos** fundamentan buena parte de la planificación de medios: de ellos se generan ratios con los que habitualmente se comparan campañas alternativas (cobertura, frecuencia, curva de respuesta, eficacia, etc). Pero su cálculo no es inmediato, como los GRP's, los costes por mil, etc; al contrario requieren algoritmos que miden complejas interrelaciones, para estimar el nº total de impactados y el nº de contactos recibidos por cada individuo. Esos algoritmos son precisamente el objeto de estudio que, por tanto, no es una investigación global de la *planificación de medios*.

b) Los programas informáticos actualmente en uso emplean esos algoritmos para calcular el alcance y la distribución de contactos. Es un comentario muy generalizado que los usuarios carecen de una adecuada formación sobre los fundamentos de los programas, que sería conveniente por varios motivos: [1] el conocimiento profundo ayuda a ponderar las salidas del ordenador, sin infravalorar su rigor ni sobrevalorar sus limitaciones; [2] teniendo presente que no existe ni solución ni algoritmo únicos, se puede demandar el algoritmo más preciso, aunque en esta decisión influyan otros factores comerciales, del entorno, etc; [3] la investigación autóctona debería romper la dependencia de la tecnología informática extranjera.

c) El volumen de **inversión publicitaria** en medios convencionales (**I Parte, §A**) requiere una planificación de medios apropiada; todo anunciante debería utilizar los procedimientos existentes y disponibles para seleccionar los medios y soportes; lo exige la racionalización de las decisiones, pero lo impiden muchas "razones", igual que no todo consumidor compra según criterios racionales. A propósito generalizo a *todos* los anunciantes, grandes, medianos y pequeños, aunque cada uno disponga de instrumentos proporcionados a su nivel; hemos distinguido **tres estadios** en el cálculo del alcance y la distribución de contactos, en función de la disponibilidad de información y del tratamiento de los datos:

1. Sobre información de estudios periódicos, hacer un tratamiento agregado de los datos. Tiene costes casi nulos, requiriendo sólo instrucción y tiempo; no es imprescindible el ordenador; producen una estimación del alcance y distribución de contactos; son los llamados *modelos de fórmula* (II §A-B; III §A-B; IV §A). Es el estadio inferior, que conviene a quienes no alcanzan otros: se aconseja pues a los pequeños usuarios, cuyo volumen de inversión no justifica mayores desembolsos.

2. Sobre información de estudios periódicos, hacer un tratamiento individual de los datos, a través de ordenadores. Se generan costes de utilización de los programas y de acceso obligatorio a la base de datos; también utilizan algoritmos para estimar el alcance y la distribución de contactos; se denominan *modelos de simulación* (III §C y IV §B). Permiten definir el *target* individuo a individuo y aporta gran velocidad al proceso de cálculo; los programas contienen diversas utilidades que ayudan a planificar de una forma más integral. Es un estadio intermedio, preferible al anterior, y actualmente es empleado en España por grandes y medianos usuarios (I §C).

3. Sobre información de estudios continuos. Admitiendo un tratamiento agregado de los datos, lo propio es hacerlo individuo a individuo. Exige ordenadores de gran potencia, habitualmente centralizados y accesibles on-line. Tiene un coste elevado, pero no sólo requiere financiación personal sino que el tamaño del mercado debe justificar tales inversiones. Se trata de los *paneles de audiencia* (I Parte, §D). Hoy en día, en España existen los audímetros para Televisión, y se restringen a los grandes usuarios. Es el estadio superior donde el alcance y la distribución de contactos no necesitan ser estimados sino directamente observados; no obstante algunos programas estiman mediante la simulación individual (p.ej. TOM, Marksel, etc). Evidentemente toda estimación está sujeta a errores, que hemos pretendido evaluar en estas páginas.

Este trabajo es eminentemente cuantitativo, aunque no pretendemos dar primacía a este enfoque; se adapta mejor a nuestra Área de Investigación, de la misma forma que los aspectos cualitativos son más propios de las Facultades de Publicidad. La adecuada

combinación de ambos enfoques debe guiar toda planificación de medios, no tanto por un sabio eclecticismo como porque se necesitan mutuamente.

Al mismo tiempo el objeto formal difiere de una ciencia exacta: los contactos entre un individuo y un soporte son contactos probables; además no implican que el individuo vea o escuche el anuncio, de manera que propiamente habría que hablar de *opportunities to see* o *opportunities to hear*. Pero, a nuestro juicio, sólo una visión reductiva quitaría el carácter científico a estas formas de conocimiento. Distingamos tres niveles: [1] un fenómeno probabilístico es exactamente medido por la Matemática Pura; [2] esos fundamentos se pueden aplicar en la medición de un fenómeno físico (por ejemplo, el impacto de un proyectil) porque influyen variables difícilmente controlables (el viento, la cantidad exacta de pólvora, etc); [3] igualmente se aplican al fenómeno social de la exposición de medios, que está influido por variables no controlables (las decisiones humanas). El carácter científico no es exclusivo del primer nivel, sino que está compartido con las ciencias aplicadas, siempre y cuando expliquen efectivamente la realidad, requisito más asequible en el segundo que en el tercer nivel.

Se trata pues de verificar la adecuación entre la realidad y el modelo que, en definición de Sissors y Bumba⁶, es "una representación simplificada de la realidad o, en el caso de la planificación de medios, una descripción de un proceso". Aquí se recogen muchos modelos -no todos- que calculan el alcance y la distribución de contactos, exclusivamente o como parte de proceso más complejo. Para testarlos seguimos el procedimiento normal expuesto por Gensch⁷: primero comprobar la lógica de su estructura interna (verificación interna) y, segundo, confrontar las estimaciones del modelo con las directas observaciones (verificación externa). En esta tesis se atiende especialmente a la validez externa, menos explorada que la interna, pero ambas tienen

⁶ SISSORS, Jack Z. and BUMBA, Lincoln: "Advertising Media Planning". NTC Business Books. NTC Publishing. 3ª Edition. 1990. Pág. 360.

⁷ GENSCHE, Dennis H. "A Computer Simulation Model for Selecting Advertising Schedules". *Journal of Advertising Research*. Vol. 6. Mayo 1969. Págs. 204 y 211.

sitio en estas páginas. La ausencia de verificaciones empíricas fue duramente criticada por Chandon⁸, sobre todo en los modelos de simulación: de ahí que en un experimento evaluáramos el modelo de simulación más utilizado en España (**I Parte, §D**). En otro estudio se analiza el alcance y la distribución de contactos estimados por un grupo de modelos de fórmula (**II Parte, §A y B**). Son pocas las cuestiones teóricas que quedan abiertas en los modelos de planificación⁹. Sin embargo, los usos y prácticas pueden ser mejorados, en todos los niveles, y ser adaptados a las fuentes de datos y programas de nuestro entorno.

La tesis se organiza en **cuatro partes**: la primera incluye varias cifras relevantes de la inversión publicitaria, la fuente de datos y los programas informáticos más utilizados en España para medios impresos y radio, exceptuando la audimetría de Televisión; también se contrastan las estimaciones de varios modelos frente a un panel. Las partes segunda, tercera y cuarta, recogen de forma sistemática los modelos multi-soportes, multi-periodos y ambos a la vez: en la segunda parte, los fundamentos teóricos se acompañan con la contrastación entre las estimaciones de tales modelos y las de los programas TOM sobre datos del EGM. Finalmente aparecen las conclusiones. Los anexos están en el diskette adjunto, para facilitar el tratamiento de los datos y reducir el papel impreso. Las Figuras y Tablas se reconocen por dos números: el primero indica la parte correspondiente, y el otro sigue la numeración dentro de esa parte.

⁸ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 561.

⁹ KILLION, Kevin C. "Media Models: Questions Needing Answers for Researchers Needing Questions". *Journal of Advertising Research*. Research Currents. Vol. 30. Nº 5. Octubre/Noviembre de 1990. Págs. RC-7 y 8.

PARTE I

PRÁCTICA **ACTUAL DE LA** **PLANIFICACIÓN**

A) INVERSIÓN PUBLICITARIA

La relevancia de la *planificación de medios* viene indicada por la inversión publicitaria en España. Se recogen aquí los últimos datos publicados de Nielsen/Repress¹⁰, la fuente más reconocida. En 1993, el volumen total del mercado publicitario "conocido" (como los mismos autores califican) se estimaba en 539.098 millones de ptas, casi un 0'9% del PIB. En la **Figura 1.2**, se desagrega esa magnitud entre los distintos Medios: Televisión, Diarios, Revistas, Suplementos + Dominicales, Radio, Exterior (Carteleras o Vallas, Cabinas Telefónicas, Marquesinas y Oppis) y Cine.

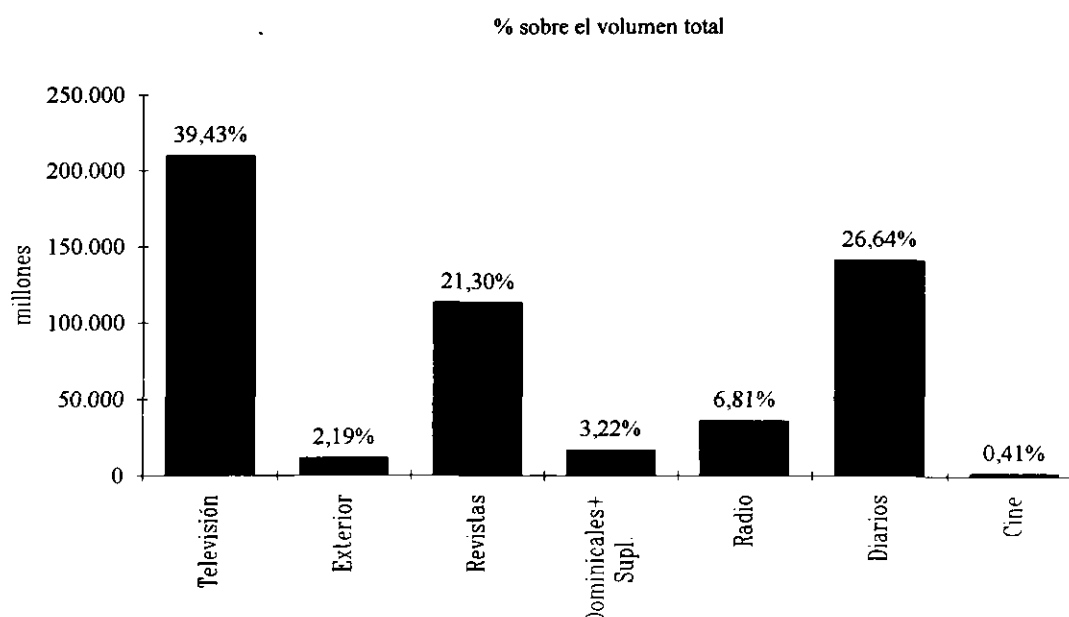


Figura 1.2: Distribución de la inversión publicitaria por medios, en España (1993).

Fuente: Nielsen /Repress (op.cit.)

En la **Tabla 1.2**, se comparan los distintos medios. Los datos de *Televisión* se basan en la cifras declaradas por las propias cadenas, debido a la gran diferencia entre las tarifas oficiales y las tarifas reales (con una relación media de 4 a 1). En *Diarios* se invirtieron 141.987 millones, correspondiendo un 71'4% a publicidad controlada (de marca y local distribuida) y el restante 28,6% a publicidad no controlada (clasificada y

¹⁰ NIELSEN/REPRESS. "Anuario de la Publicidad: 1994". Ed. A. C. Nielsen Company, S.A. Madrid.

local no distribuida). Los títulos medidos son 86, aproximadamente el doble que los estudiados por el EGM. La inversión de *Revistas* es de 113.523 millones, de los cuales sólo un 43% corresponden a las 121 publicaciones controladas; del resto de publicaciones (con un contenido especializado o de menor importancia) no se presenta más información, estimando su inversión en el restante 57%. La *saturación media* se define como el n° de minutos/páginas de publicidad entre el n° de minutos/páginas totales emitidos o publicados.

MEDIO	Inversión Total (millones)	N° de Cadenas/ Soportes	N° de marcas	N° de anunciantes	N° de anuncios	Saturación media (%)
Televisión	210.145	13	4.976	2.985	776.707	4,7
Diarios	141.987	86	23.330	16.337	707.967	12,3
Supl.+Dom.	17.144	34	2.705	1.673	15.453	19,3
Revistas	113.523	121	9.069	5.783	74.656	21,3
Radio	36.291	-	3.166	3.959	1.670.721	-

Tabla 1.2: Mercado publicitario por los principales medios. Fuente: Nielsen...(op.cit.).

De todos estos datos se pueden extraer algunas **conclusiones** que afectan a nuestro objeto de estudio:

1. Supongamos por un momento que todos los anuncios en Diarios, Revistas, Suplementos + Dominicales y Radio, se han planificado con los programas TOM. Es obvio decir que entonces los modelos de fórmula no tendrían ninguna aplicación práctica. Pero me pregunto: ¿no se podría planificar mejor los 40.000 millones de la publicidad no controlada en Diarios (clasificada y local no distribuida) o quizá hay que suponer también que esos anunciantes utilizan los TOM?; ¿y los 65.000 millones de las revistas no controladas o los 7.500 millones de la publicidad local no controlada en Radio? Al mismo tiempo los modelos de fórmula se aplican sobre datos de audiencia que, como luego comentaré, no existen de todas las publicaciones ni cadenas: en esos casos no cabe planificación en función del alcance y la distribución de contactos.

3. ¿Quién puede creer que todos los anuncios controlados en Diarios y Revistas se han planificado con los programas TOM? El análisis de la concentración de la inversión publicitaria (**Tabla 1.3**) aclarará definitivamente esta cuestión. No obstante, entiéndanse los siguientes juicios como lo que son: generalizaciones, realizadas además sobre unos valores medios. Parece lógico que los 250 primeros anunciantes, que gastan de media entre 100 y 300 millones, utilicen las mejores técnicas de planificación del mercado, en manos de las agencias y centrales contratadas. Pero no se puede decir lo mismo del resto de anunciantes, 16.087 en Diarios y 8.819 en Revistas, que apenas gastan entre 1 y 4 millones por término medio. Quizá normalmente no se planifique a esos niveles: ¿pero supondría algún coste adicional planificar con los modelos de fórmula? Por contra, si se obtiene un beneficio adicional, que ahora me resisto a explicar. Descrito el problema, convendría evaluar qué parte de la inversión no se planifica con los TOM y podría emplear procedimientos más apropiados a sus posibilidades.

	Inversión Total controlada	Inversión de los 250 primeros anunciantes	Inversión media de esos 250 anunciantes	Inversión del resto de anunciantes	Inversión media del resto de anunciantes
Diarios	101.379	73.196	292,78	28.183	1,75
Revistas	48.814	31.143	124,57	17.671	3,19

Tabla 1.3: Concentración de la inversión (en millones). Fuente: Nielsen... (op.cit.).

3. Algunos datos muestran que no todas las *agencias* pueden estar usando los modelos de simulación. Según el AIMC¹¹, en 1991 había 41 agencias asociadas y, por tanto, en condiciones de acceder al fichero individualizado. Tres años después, en España se anunciaban unas mil agencias¹², un número mucho mayor. El ranking de

¹¹ ASOCIACIÓN PARA LA INVESTIGACIÓN DE MEDIOS DE COMUNICACIÓN (AIMC). ESTUDIO GENERAL DE MEDIOS (EGM). Resultados Acumulados 1ª + 2ª + 3ª Olas 1991: Individuos y Hogares. Febrero - Noviembre de 1991.

¹² GUÍA DE MEDIOS DE COMUNICACIÓN. Nº 119. Septiembre-Noviembre de 1994. Ed: Remarca, S.A. Madrid.

facturación de las primeras 71 agencias de 1994¹³, aun con la reserva que impone la forma de obtención de esos datos, deja ver un buen grupo que, al menos en principio, no tiene conexión con el EGM y factura un volumen considerable. Hay que considerar también que la planificación se puede encargar a *centrales*, que generalmente están vinculadas al EGM y utilizan programas de simulación.

4. Cabe preguntarse si esta importante partida de la economía nacional es capaz de financiar un *panel de audiencia*. Hacemos dos consideraciones: por una parte, la facturación de los estudios de mercado ha ido creciendo en los últimos años¹⁴, aunque la objetividad de tales datos está limitada por su forma de recogida; los estudios sistemáticos (paneles, continuos y ómnibus) facturaron 5.150 millones en 1992, 8.600 millones en 1993 y 8.050 millones en 1994: aunque no sean magnitudes directamente comparables, tales estudios suponían, en el último año, escasamente el 1'6% del volumen publicitario total; estas cifras apuntan que probablemente el gasto en investigación de medios sea una pequeñísima parte de la inversión en medios; entonces el problema no radicaría exclusivamente en el tamaño del mercado. Por otra parte, habría que evaluar su rentabilidad, de manera que la información obtenida justifique los importantes costes de esos paneles; suponiendo además que todas las partes se pongan de acuerdo.

¹³ ANUNCIOS REVISTA. Nº 44. Marzo de 1995. Págs. 4 y 5.

¹⁴ ALÓS, Joan S. "Industria de los Estudios de Mercado en España 1992" (Nº 44. Marzo de 1994. Pág. 104-6); "Industria de los Estudios de Mercado en España 1993" (Nº 47. Marzo de 1995. Págs. 74-6); "Industria de los Estudios de Mercado en España 1994" (Nº 48. Julio de 1995. Págs. 74-7). *Investigación y Marketing*. AEDEMO.

B) ESTUDIO GENERAL DE MEDIOS

El Estudio General de Medios (EGM) es una organización sin ánimo de lucro, integrada en la Asociación para la Investigación de Medios de Comunicación (AIMC). En 1991, los asociados de AIMC¹⁵ eran 44 periódicos, 86 revistas, 20 emisoras de radio, 9 cadenas de televisión, 41 agencias de publicidad, 17 centrales de compra y 4 anunciantes. González Lobo¹⁶ señala que "la Asociación está organizada en dos estamentos, uno formado por los medios y otro compuesto por los restantes miembros". Esos estamentos tienen el mismo peso en las decisiones de la Asamblea General, órgano supremo de la Asociación. La Junta Directiva está formada por ocho miembros elegidos democráticamente por la Asamblea General, cuatro por cada estamento. La Comisión Técnica consta de nueve miembros, seis elegidos por la Asamblea (mitad de los medios y mitad del resto) y tres por la Junta Directiva. Según Méndez¹⁷, "este cuidadoso entramado de relaciones y equilibrios [...] garantiza la objetividad y neutralidad de la investigación". González Lobo también indica que la financiación de los estudios proviene de las cuotas de los miembros; y del total corresponde un 75% a los medios (con cuotas proporcionales a su audiencia) y un 25% a las partes restantes. Acerca de la evolución histórica del EGM, puede consultarse el trabajo de Mazo¹⁸.

El Estudio General de Medios es un estudio *multimedia-producto*, es decir, investiga simultáneamente la audiencia de varios *medios* (televisión, prensa diaria y semanal, revistas, radio y cine) y el consumo de *productos*, así como el equipamiento del hogar y los estilos de vida de los entrevistados. La investigación se realiza mediante

¹⁵ ASOCIACIÓN PARA LA INVESTIGACIÓN DE MEDIOS DE COMUNICACIÓN (AIMC). ESTUDIO GENERAL DE MEDIOS (EGM). Resultados Acumulados 1ª + 2ª + 3ª Olas 1991: Individuos y Hogares. Febrero - Noviembre de 1991.

¹⁶ GONZÁLEZ LOBO, Mª Ángeles. "Curso de Publicidad". Eresma & Celeste Ediciones. Madrid. 1994. Pág. 266.

¹⁷ MÉNDEZ RODRÍGUEZ, JUAN LUIS. "Método del recuerdo. Estudio General de Medios". Ponencia presentada en el 1º Seminario de Audiencia sobre Radio (49º Seminario de AEDEMO). Madrid. Noviembre 1989.

¹⁸ MAZO DEL CASTILLO, Juan Manuel. "Las decisiones sobre medios en comunicación por objetivos". Tesis doctoral, dirigida por Dr. D. Justo Villafañe. Dpto. de Comunicación Audiovisual y Publicidad I. Universidad Complutense de Madrid. Madrid, 1991. Págs. 553-93.

entrevistas personales en el hogar. La muestra es de unas 44.000 entrevistas anuales, distribuidas en tres olas de 14.666 aproximadamente, cuyo trabajo de campo se prolonga durante un mínimo de 7 semanas y un máximo de 9 semanas, en estos meses: febrero-marzo (1ª ola), mayo-junio (2ª ola) y octubre-noviembre (3ª ola). Méndez, Director Gerente de AIMC, exponía que "no es un estudio continuo, pero tiene una longitud y distribución de campo suficientemente amplia para recoger adecuadamente las variaciones estacionales en el consumo de los medios"¹⁹. Acerca de la evolución del cuestionario, puede consultarse la ponencia de Clemente²⁰.

La muestra se distribuye geográficamente en dos fases: primero, la mayor parte de la muestra se adjudica proporcionalmente a la población de cada provincia; después, el resto se asigna a algunas provincias con televisiones autonómicas, diarios locales, etc. Entonces, debido a la estructura geográfica de los medios, cada individuo muestral no representa el mismo número de individuos reales. Dentro de las provincias, las entrevistas se reparten en función del tipo de hábitat, medido según el tamaño del municipio. Para llegar a toda la población, incluso la diseminada, Méndez indica que sobre la sección censal se disponen las direcciones de partida de las *rutas aleatorias*, que conducen hasta el edificio y vivienda y, una vez dentro del hogar, el entrevistado se elige mediante una tabla de números aleatorios. En ese momento, el promedio de entrevistas por ruta aleatoria era de 6, existiendo 1.090 puntos de muestreo por ola y 2.196 en todo el año.

La información de los medios obtenida de las encuestas es variada: para todos, audiencia del último periodo y hábito de audiencia; además, para medios impresos, forma de abastecimiento (suscripción, compra personal,...), destinatario principal de la publicación y cantidad de lectura; para radio, horario y lugar de escucha; para televisión,

¹⁹ MÉNDEZ RODRÍGUEZ, JUAN LUIS. "La medición de la audiencia de radio y sus alternativas". Audiencia y Programación. RTVV. Valencia. 1993. Pág. 111.

²⁰ CLEMENTE, Jorge. "El Estudio General de Medios (I)". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios-Televisión. Gandía. Febrero 1985. Págs. 121-48.

canales sintonizados y vídeo (del hogar y comunitario), calidad de imagen, preferencias de programas, etc. La audiencia de los medios se considera la *audiencia del último periodo*, y se recoge en cuatro informes anuales (tres olas y el acumulado) que AIMC publica para sus asociados; esos datos son globales para todos los individuos y para las amas de casa, y se cruzan con algunas variables de clasificación, de consumo, etc; aunque sea información de gran valor, a veces es insuficiente porque no se adapta al *target* requerido, no cruzan todas las variables, etc.

Dos son las utilidades del EGM, según González Lobo: "una es conocer de forma global la audiencia y estructura general de los medios y de los principales soportes, establecer comparaciones y analizar tendencias; y otra, servir de base de datos para la planificación de medios"²¹. La planificación, objeto de estudio de esta tesis, requiere un análisis más profundo de los medios y consumidores; se consigue con el tratamiento informático de los datos del EGM, a través de los programas TOM del Centro de Cálculo ODEC: así se aprovecha al máximo la información, pudiendo definir cualquier *target* (que se adapte a las respuestas del cuestionario), cruzar todas las variables, etc. Para ello hay que definir la probabilidad que tiene cada individuo ante cada soporte: ese proceso, llamado probabilización, se determina a partir del *hábito de audiencia* y de la *audiencia del último periodo* (véase **III Parte, § C. 5**). Entonces, a través de la simulación, se puede determinar el probable comportamiento individual ante uno o varios soportes y, agregando todos los individuos, el comportamiento global del *target*.

Se generan entonces ligeras diferencias entre la audiencia del último periodo (recogida en los informes) y la audiencia probabilizada o probable audiencia (utilizada en la planificación): en la **Tabla 1.4** se muestra que las divergencias son pequeñas, por ejemplo, en las cinco primeras revistas utilizadas en nuestra investigación.

²¹ GONZÁLEZ LOBO, M^a Ángeles. "Curso de Publicidad". Eresma & Celeste Ediciones. Madrid. 1994. Pág. 281.

REVISTAS	Audiencia último periodo (.000)	Audiencia probabilizada (.000)
Teleprograma	4.641	4.633
Teleindiscreta	4.092	4.090
Pronto	3.692	3.685
Muy Interesante	2.251	2.252
Hola	2.410	2.404

Tabla 1.4

La investigación realizada en las **Partes II** y **IV** se ha hecho sobre el acumulado de las tres olas de 1991, cuya ficha técnica se expone en la **Figura 1.3**. El universo, el diseño muestral y el trabajo de campo, actualmente se definen de la misma manera, *mutatis mutandis*.

UNIVERSO

Individuos mayores de 14 años, residentes en hogares unifamiliares, ubicados en municipios dentro de la España Peninsular, Islas Baleares e Islas Canarias.

DISEÑO MUESTRAL

Polietápico estratificado, básicamente proporcional:

- por días de la semana.
- por tipo de hábitat (tamaño de municipio).
- por provincia.

Tamaño muestral: 43.860 entrevistas reales.

TRABAJO DE CAMPO**Selección de unidades muestrales:**

- Selección del hogar, mediante el procedimiento de las rutas aleatorias.
- Selección de la persona (sólo una) en el hogar, mediante tabla de números aleatorios.
- Sustituciones por personas del mismo sexo, categoría socio-económica e intervalo de edad.

Plazo de realización:

1ª Ola: 31 de enero - 9 de abril de 1991

2ª Ola: 3 de mayo - 4 de julio de 1991

3ª Ola: 8 de octubre - 3 de diciembre de 1991

Supervisión:

- un 17% realizada por supervisores de E.C.O. mediante contraentrevista personal.
- un 10% realizada directamente por AIMC.

Proceso de datos: realizado por ODEC.

Ampliación Muestral en la Comunidad Valenciana.

Figura 1.3: Ficha técnica de 1991. Fuente: AIMC-EGM²².

La variable *clase social* se explica en función de la ocupación y del nivel de instrucción del cabeza de familia (aquél del cual provienen los ingresos fundamentales de la familia). Puede tomar cinco alternativas: alta, media-alta, media-media, media-baja y baja. A cada combinación de ocupación y nivel de instrucción le corresponde un tipo de clase social, según la **Tabla 1.5** (nótese que el nivel de instrucción se refiere a estudios terminados).

²² ASOCIACIÓN PARA LA INVESTIGACIÓN DE MEDIOS DE COMUNICACIÓN (AIMC). ESTUDIO GENERAL DE MEDIOS (EGM). Resultados Acumulados 1ª + 2ª + 3ª Olas 1991: Individuos y Hogares. Febrero - Noviembre de 1991.

OCUPACIÓN	NIVEL DE INSTRUCCIÓN								
	No sabe leer	Sabe leer	Primar. incompletos	1º grado	2º grado 1º ciclo	2º grado 2º ciclo	3º grado med.	3º grado sup.	N.S./N.C.
Agr. Cooperativa	E	E	E	E	D	C	C	C	D
Agr. sin empleados	E	E	E	E	D	C	C	C	D
Agr. 1-5 empleados	D	D	C	C	C	B	B	B	C
Agr. 6/+ empleados	C	C	C	C	C	B	B	A	B
Cte. 1-5 empleados	C	C	C	C	B	B	A	A	B
Cte. sin empleados	D	C	C	C	B	B	B	B	C
Cte. 6/+ empleados	C	C	B	B	A	A	A	A	B
Profesional liberal	E	D	C	C	B	B	A	A	C
Trabajador manual	E	E	D	D	C	C	C	C	C
Director gran empres.	C	B	B	B	A	A	A	A	A
Director peq. empres.	D	D	C	C	B	B	A	A	C
Mando superior	D	C	C	B	B	B	A	A	C
Mando intermedio	D	D	C	C	C	B	B	B	C
Capataces	E	E	D	D	C	C	B	B	C
Representante	E	D	D	C	C	C	B	B	C
Administrativo	E	D	C	C	C	C	B	B	C
Obrero especializado	E	E	D	D	C	C	C	C	C
Vendedores	E	D	C	C	C	C	C	B	C
Obreros no especial.	E	E	E	D	D	D	C	C	D
Subalternos	E	E	E	D	D	D	C	C	D
Otro no cualificado	E	E	E	D	D	D	C	C	D
Jornaleros	E	E	E	D	D	D	C	C	D
RESTO	E	E	E	D	D	C	C	C	-

Tabla 1.5: Clases sociales: A (alta), B (media-alta), C (media-media), D (media-baja) y E (baja). Fuente: AIMC-EGM²³.

Los márgenes de confianza o error²⁴ se determinan según los distintos niveles de audiencia; el porcentaje de error e se expresa en función del porcentaje de audiencia p , de su complementario q y del tamaño de la muestra N :

$$e (\%) = \pm 2 \cdot \sqrt{\frac{p \cdot q}{N}}$$

²³ ASOCIACIÓN PARA LA INVESTIGACIÓN DE MEDIOS DE COMUNICACIÓN (AIMC). ESTUDIO GENERAL DE MEDIOS (EGM). Resultados Acumulados 1ª + 2ª + 3ª Olas 1991: Individuos y Hogares. Febrero - Noviembre de 1991.

²⁴ LOZANO CHOZA, Domingo. "El Estudio General de Medios (II)". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios-Televisión. Gandía. Febrero 1985. Págs. 150-1.

Dada la estructura de la fórmula, el error sería el mismo con valores de p complementarios (p. ej. 20% y 80%). Por eso, Lozano indica que es mejor utilizar el coeficiente de variación o error relativo, que tiene en cuenta la magnitud de la audiencia:

$$CV (\%) = \pm \frac{2 \cdot \sqrt{\frac{p \cdot q}{N}}}{p}$$

Veamos un ejemplo: Teleprograma, durante 1991, tenía una audiencia total de 4.630.616, estimada con una muestra de 43.860 entrevistas, que representan a los 32 millones de individuos mayores de 14 años. Entonces la tasa de audiencia p vale 14'47%, y su complementario q es igual a 85'53%. Resulta un margen de error absoluto de $\pm 0'34\%$ de la población (± 107.500 individuos en números absolutos); por tanto, el intervalo de confianza de la tasa de audiencia es 14'13-14'81% (el intervalo de la audiencia absoluta varía entre 4.523.116 y 4.738.116). Por otra parte, el coeficiente de variación del $\pm 2'32\%$ (± 742.950 individuos).

El planificador debe vigilar el coeficiente de variación durante el proceso y tomar las decisiones tras evaluar los márgenes de confianza de la información. Así, cuando ha definido el *target*, puede determinar el coeficiente de variación de cada soporte, a partir de su correspondiente audiencia útil y del número de entrevistas dentro de ese público objetivo. En palabras de Lozano, "el EGM, como todo instrumento, tiene sus limitaciones y, por tanto, no debe ser utilizado por encima de sus posibilidades. [...] Si además de aumentar las posibilidades de error, añadimos los sofisticados sistemas de probabilización, coeficientes de memorización, etc, podemos estar dando resultados que nada tengan que ver con la realidad".

C) PROGRAMAS DE PLANIFICACIÓN TOM

Pistoni²⁵ propone que el proceso lógico de planificación, a través de los programas disponibles en ODEC²⁶, debería seguir estas etapas:

1. Definición del *target* y análisis del producto/servicio y de sus consumidores. Los programas disponibles son: TOMSEG (segmentación), TOMPER (perfiles socio-demográficos), TOM0 (cuantificación del público objetivo), TOM5 (tabulación; cruce de variables) y ODIN (análisis general de variables).

2. Selección del medio o combinación de medios. Cuenta con los programas TOMEST (estructura de medios), TOMFAC (análisis factorial de correspondencias) y ODIN (análisis general de variables).

3. Selección de soportes pertenecientes a los medios elegidos: TOM1 y Atelier RTV (clasificación de soportes; ranking), TOM1D (duplicación entre soportes), TOM1ES y TOM1EP (análisis de la eficacia de los soportes/productos), TOMCOB (análisis de la saturación de las coberturas) y ODIN (análisis general de variables).

4. Construcción de planes para rentabilizar al máximo el presupuesto combinando los soportes anteriormente seleccionados, y para alcanzar los objetivos de cobertura/frecuencia: TOM3 (optimización de planes), TOM4 (análisis de alternativas), Atelier RTV (construcción de planes con resultado en tiempo real) y ODIN (optimización de planes).

5. Evaluación de la campaña, obteniendo los resultados que aportarían los diferentes planes construidos, valorando las alternativas y distribuyendo los planes en el tiempo: TOM2 (evaluación de planes), TOM2M (análisis de la aportación de cada

²⁵ PISTONI DURÁN, Luis. "El tratamiento de la información para la planificación de medios". Ponencia recogida en: Audiencia y Programación. Editado por el Ente Público RTVV. Valencia. 1993.

²⁶ Actualmente el autor ocupa el cargo de Director Comercial de ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A.

medio a la campaña), TOM2C/CX (distribución de los planes en el periodo de la campaña) y Atelier RTV (análisis y descripción de los planes).

El programa Atelier RTV, según la información facilitada directamente en ODEC, se ha desechado en la fecha de lectura de esta tesis. Los programas actualmente disponibles son los TOM y el ODIN pero este último es menos importante, de manera que nos centraremos en los TOM.

El paquete informático TOM fue realizado por Canguilhem para el Cabinet de Modelisation par Calculateur (CMC) de París. Según Pistoni²⁷, en 1983 fue adaptado y puesto en marcha en España por UNITEC, S.A., actualmente ODEC. Su sistema conversacional permite al usuario modificar las decisiones durante el proceso. Hasta 1993 era de uso exclusivo en Mainframe, accediendo on-line a los ordenadores del Centro de Cálculo ODEC. Después se ha mejorado con el Interface TOM-Micro, que permite su uso en ordenadores personales, con un microprocesador al menos del tipo 486: con ello se evitan todos los problemas del acceso on-line (disponibilidad de línea telefónica, disponibilidad de ordenador, etc); se ha facilitado mucho la petición de los datos, con un sistema de ventanas que evita los anteriores lenguajes codificados; se evita el divorcio entre tarifas oficiales y reales, puesto que ya no se paga por salida de ordenador, sino por el derecho de uso de los programas. Como desventaja, no todos los programas están disponibles actualmente en el TOM-Micro, y muchas empresas emplean simultáneamente el acceso on-line. Como predecía Canguilhem²⁸ cuando se iniciaban los ordenadores personales, su aplicación a la planificación "supondría una ganancia en productividad para los planificadores: entrar menos, explotar más", no como una alternativa superadora sino como un complemento más flexible.

²⁷ PISTONI DURÁN, LUIS. "Programas actuales". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Pág. 253.

²⁸ CANGUILHEM, Jean François. "Futuro de la informática aplicada a la planificación". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Pág. 272.

Voy a exponer el paquete informático TOM que ofrece la empresa ODEC. Se pretende tanto su descripción comercial²⁹ como el análisis de sus fundamentos teóricos. Se ha de tener en cuenta que estos programas están en constante actualización, y todo lo dicho se refiere al momento de lectura de esta tesis.

²⁹ Recogida de los manuales ofrecidos por la empresa:

- TOM: Programas para la Planificación de Medios. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A.
- TOM-MICRO: Manual de Referencia. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A. 1994.
- TOM-MICRO: Generación de Resultados. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A. 1995.

1. Programas disponibles en Mainframe y en Tom-Micro

TOM0 CUANTIFICACIÓN DE PÚBLICOS OBJETIVOS

Este programa permite cuantificar cualquier público objetivo, sobre el fichero que interese (p.ej. primera ola de 1995 o primer acumulado de 1995). El público objetivo se puede definir en función de cualquier variable o grupo de variables socio-demográficas, de consumo personal, de consumo individual, de equipamiento, de posesión de animales, de consumo en el hogar (amas de casa) y de estilos de vida. Permite construir un *target* formado simultáneamente por varios segmentos. Se puede ponderar específicamente cada segmento, cada variable y cada alternativa. La salida de ordenador ofrece el público objetivo real (que pondera todas las categorías con valor 100) y el público objetivo ponderado (cuando los ítems se ponderan con distintos valores). También ofrece el número de casos o encuestas de la muestra, que deberá ser revisado para evitar errores muestrales (véase el coeficiente de variación, en el epígrafe del EGM). La especificación del fichero y la definición del público objetivo es un input obligado de todos los TOM, y estos resultados se presenta como cabecera de todas las salidas.

TOM1 CLASIFICACIÓN DE SOPORTES

Como en todos los programas, hay que indicar el fichero de uso y definir el público objetivo; además, en este caso, la lista de soportes a clasificar y el tipo de precio; para simplificar se puede seleccionar una/s familia/s, que agrupa/n varios soportes. La información disponible incluye todos los soportes asociados al AIMC y, por tanto, aquellos investigados en las entrevistas del EGM. En cuanto al precio, se puede elegir entre cuatro tipos estándar según las características de los medios, como aparece en la **Tabla 1.6**. La base de datos incluye las tarifas oficiales de todos los soportes con esas configuraciones; pero el programa permite introducir libremente las tarifas reales que el planificador consiga después de descuentos; además existe tal variedad de tarifas, que sería casi imposible clasificarlas para todos los soportes.

	Tipo 1	Tipo 2	Tipo 3	Tipo 4
TELEVISIÓN	20 segundos	25 segundos	30 segundos	45 segundos
PRENSA	pág. b/n	media pág. b/n	doble pág. b/n	media doble pág. b/n
REVISTAS/ SUPLEMENTOS	pág. color	media pág. color	doble pág. color	media doble pág. color

Tabla 1.6

La salida incluye un ranking de soportes, ordenados de mayor a menor según varios criterios: secuencial (orden alfabético), cobertura (contactos en el *target* o útiles), rentabilidad (coste por mil contactos útiles) y afinidad (contactos útiles dividido por contactos totales del soporte). En el estudio propuesto en la **II Parte, § A.1.4**, se detalla la selección de soportes según estos criterios. La cobertura y la rentabilidad se ordenan respectivamente según los Rating Point (contactos útiles dividido entre público objetivo ponderado) y coste por Rating Point (precio dividido entre Rating Point). Cuenta con una ayuda gráfica, que representa la posición de cada soporte según dos ejes, la cobertura en abcisas y la rentabilidad en ordenadas.

TOMID DUPLICACIÓN ENTRE SOPORTES

Proporciona las duplicaciones de audiencia entre dos soportes cualesquiera, de todos los medios. Se solicita un soporte cabecera y unos soportes filas, y se obtiene la duplicación del primero con los demás. La información se presenta en valores absolutos y en porcentaje al público objetivo ponderado. Ofrece el porcentaje que supone la duplicación entre la audiencia neta de los soportes implicados. También ofrece el índice de discriminación, cuyo valor oscila entre 100 cuando la duplicación es total, y -100 cuando no existe duplicación. Canguilhem empleó este concepto análogamente para definir el público objetivo³⁰. La **discriminación** mide el poder explicativo de la audiencia de un soporte respecto a la de otro, así como la proximidad entre ambas. Para ello, hay que analizar tanto la proximidad entre las audiencias, como el alejamiento de sus complementos (las no audiencias). La proximidad o alejamiento se mide a través del concepto de distancia. La discriminación es la diferencia entre la distancia del primer soporte al segundo y la distancia del primero al complemento del segundo. En la **Figura 1.4**, se observa gráficamente el concepto de distancia entre dos soportes.

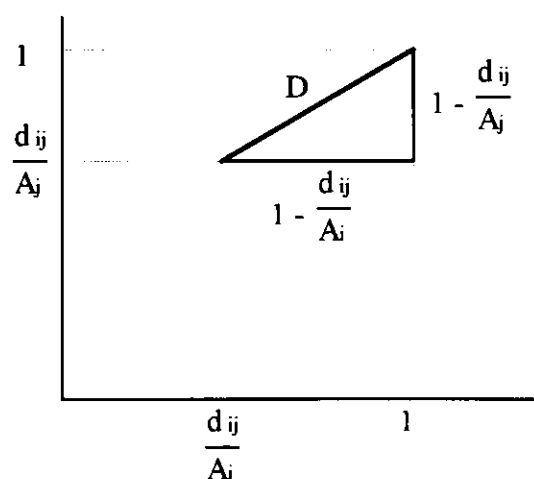


Figura 1.4: Distancia D entre las partes duplicadas de los soportes i y j .

³⁰ CANGUILHEM, J. F. "Cómo definir prácticamente una población-objetivo publicitaria". *Información de publicidad y marketing*. Nº 119. 1973. Págs. 37-44.

Aplicando el teorema de Pitágoras, y normalizando por la distancia máxima $\sqrt{2}$, se calcula el valor de D:

$$D\left(\frac{d_{ij}}{A_i}, \frac{d_{ij}}{A_j}\right) = \frac{1}{\sqrt{2}} \cdot \sqrt{\left(1 - \frac{d_{ij}}{A_i}\right)^2 + \left(1 - \frac{d_{ij}}{A_j}\right)^2}$$

El anterior valor varía entre 0 y 1 (0 y 100 en porcentajes), tomando el valor 0 cuando la duplicación es nula (punto de máxima concomitancia) y el valor 1 cuando las audiencias se identifican. De manera similar se calcula:

$$D\left(\frac{A_j - d_{ij}}{A_j}, \frac{A_j - d_{ij}}{P - A_i}\right) = \frac{1}{\sqrt{2}} \cdot \sqrt{\left(1 - \frac{A_j - d_{ij}}{A_j}\right)^2 + \left(1 - \frac{A_j - d_{ij}}{P - A_i}\right)^2}$$

El índice de discriminación se define como la diferencia de las dos distancias anteriores:

$$ID_{ij} = \frac{1}{\sqrt{2}} \cdot \sqrt{\left(1 - \frac{A_j - d_{ij}}{A_j}\right)^2 + \left(1 - \frac{A_j - d_{ij}}{P - A_i}\right)^2} - \frac{1}{\sqrt{2}} \cdot \sqrt{\left(1 - \frac{d_{ij}}{A_i}\right)^2 + \left(1 - \frac{d_{ij}}{A_j}\right)^2}$$

Como no cumple la propiedad conmutativa (el coeficiente de i sobre j no es igual que el de j sobre i), el TOMID calcula el índice de discriminación como la media aritmética de ambos casos:

$$ID = \frac{1}{2 \cdot \sqrt{2}} \cdot \sqrt{\left(1 - \frac{A_j - d_{ij}}{A_j}\right)^2 + \left(1 - \frac{A_j - d_{ij}}{P - A_i}\right)^2} + \frac{1}{2 \cdot \sqrt{2}} \cdot \sqrt{\left(1 - \frac{A_i - d_{ij}}{A_i}\right)^2 + \left(1 - \frac{A_i - d_{ij}}{P - A_j}\right)^2} - \frac{1}{\sqrt{2}} \cdot \sqrt{\left(1 - \frac{d_{ij}}{A_i}\right)^2 + \left(1 - \frac{d_{ij}}{A_j}\right)^2}$$

TOM2 EVALUACIÓN DE PLANES

Hay que suministrarle los soportes seleccionados, con sus ponderaciones (media-weight) y las inserciones en cada soporte. Permite desglosar los resultados según cualquiera de las variables, y evaluar a la vez varios planes alternativos. De cada plan evalúa las siguientes magnitudes:

- Alcance neto en miles de individuos y cobertura en porcentaje sobre el público objetivo.
- Coste por mil impactos: el presupuesto dividido por los miles de contactos.
- Alcance bruto en miles de individuos.
- Coste por mil impactados: el presupuesto dividido por los miles de contactados.
- GRP's (Gross Rating Points): el cociente entre la audiencia bruta y el público objetivo (en términos porcentuales); o bien el producto entre la cobertura y la frecuencia; o también la suma de los Ratings de las inserciones incluidas.
- OTS (Opportunities To See): cociente entre la audiencia bruta y la audiencia neta, o la media de contactos recibidos; más correctamente debería llamarse frecuencia.
- Presupuesto: suma de los precios de las inserciones seleccionadas.
- Distribución de contactos: se expresa en porcentajes sobre el público objetivo; la salida muestra los que han recibido n contactos, y los que han recibido n o más contactos. Una opción admite la representación gráfica. También se puede ponderar la curva de respuesta, asignando diferentes pesos a los niveles de la distribución de contactos.

Todas estas magnitudes se calculan inmediatamente a partir del alcance neto acumulado y de la distribución de contactos, que es lo difícil de medir; para ello se han planteado muchas alternativas, que tendremos ocasión de estudiar en las **Partes II, III y**

IV. Ahora se explica cómo los calcula el TOM, describiendo el proceso desde la recogida de datos.

Proceso de estimación del alcance y la distribución de contactos

1. Medición de la audiencia del último periodo y de los hábitos de audiencia, a partir de entrevistas personales del Estudio General de Medios. Acotando al medio revistas semanales, la información proviene de las preguntas de la **Figura 1.5**.

1. A continuación le voy a enseñar el título de unas determinadas revistas (se muestra cartón con títulos). Le agradecería que me fuera diciendo cuáles de ellas ha leído u hojeado en los últimos 6 meses.

2. De cada cuatro ejemplares de ... (uno de los anteriores) que salen al mes, cuantos acostumbra a leer u hojear?

- 4 sobre 4.
- 3 sobre 4.
- 2 sobre 4.
- 1 sobre 4.
- Menos de 1 sobre 4.
- Ninguno.

3. ¿Cuándo ha leído u hojeado el último ejemplar de ...?

- Últimos 7 días.
- De 8 a 15 días.
- De 15 días a 1 mes.
- De 1 mes a 2 meses.
- Más de dos meses.

Figura 1.5

2. Los hábitos de exposición (*preg. 2*) y la última exposición (*preg. 3*) han de ser probabilizados, de manera que al individuo se le asigne una *probabilidad de exposición* ante cada soporte (según se explica en **III Parte, § C.5**). Se hace la hipótesis de que esa probabilidad se mantiene constante en el tiempo. Como resultado, tendremos la **Tabla 1.7** que cruza los individuos entrevistados y la probabilidad de exposición a todos los soportes; supongamos n individuos y m soportes:

Individuos	Soportes				
	S ₁	S ₂	S ₃	...	S _m
1	P ₁₁	P ₁₂	P ₁₃		P _{1m}
2	P ₂₁	P ₂₂	P ₂₃		P _{2m}
3	P ₃₁	P ₃₂	P ₃₃		P _{3m}
...					
n	P _{n1}	P _{n2}	P _{n3}		P _{nm}

Tabla 1.7

3. Los programas TOM determinan la cobertura y la distribución de contactos de una campaña a partir del *cálculo de la simulación individual*, así llamado en contraposición a los *modelos de fórmula*. La notación no se ajusta a la realidad pues la *simulación* también se basa en fórmulas. Se diferencian en que la *simulación* estima a partir de las probabilidades individuales, mientras que los *modelos de fórmula* utilizan los datos a nivel agregado. El modelo TOM es un modelo de simulación (véase **Parte IV**) que "como su propio nombre indica, simula el comportamiento de los individuos respecto a los medios, mediante la utilización del cálculo de probabilidades, de manera que se pueda llegar a establecer el objetivo básico de cualquier modelo de planificación, es decir, el cálculo de las distribuciones de frecuencia de los individuos con las campañas de publicidad"³¹. La cobertura del individuo *i* tiene la siguiente expresión³²:

$$C_i = 1 - \prod_{j=1}^m (1 - P_{ij})^{r_j}$$

siendo P_{ij} la probabilidad de que el individuo *i* contacte con el soporte *j*, y r_j el número de inserciones del soporte *j*.

La distribución de contactos se determina, para cada individuo, a través de G_i :

³¹ INSTITUTO NACIONAL DE PUBLICIDAD - EMOPÚBLICA: "Investigación sobre Técnicas de Selección y Planificación de Medios Publicitarios, II: Recopilación de Modelos de Planificación de Medios". Madrid. Mayo de 1978. Pág. 139.

³² LAMAS ALONSO, Carlos: "Eso de las evaluaciones no lo veo claro". Ponencia presentada en el 9º Seminario sobre Audiencia de Televisión (62º Seminario AEDEMO). Lisboa. Febrero 1993. Recogida bajo el mismo título en la revista:

- *Investigación y Marketing*. Asociación Española de Estudios de Mercado, Marketing y Opinión (AEDEMO). Nº 42. Julio 1993. Págs. 58-62.

$$G_i = \sum_{j=1}^m r_j \cdot P_{ij}$$

Del cociente G_i / C_i resultará una parte entera (E_i) y una parte decimal (D_i). El conjunto de personas (M_i), que representa cada individuo, se divide en tres grupos:

- 1º Grupo: formado por $(1-C_i) \cdot M_i$ personas, tiene asignados 0 contactos.
- 2º Grupo: formado por $(1-D_i) \cdot C_i \cdot M_i$ personas, tiene asignados E_i contactos.
- 3º Grupo: formado por $D_i \cdot C_i \cdot M_i$ personas, tiene asignados E_i+1 contactos.

El alcance y la distribución de contactos final de una campaña se logra sumando los resultados obtenidos para cada individuo. Veámoslo a través de un **ejemplo**: una campaña consta de 3 inserciones en el soporte S_1 , 4 en S_2 , 2 en S_3 , y 3 en S_4 . Conocemos las probabilidades de exposición de cuatro entrevistados (**Tabla 1.8**), suponiendo que cada uno representa a 750 personas (en la práctica, los individuos muestrales representan distinto nº de individuos poblacionales).

Entrevistados	Soportes			
	S_1	S_2	S_3	S_4
1	0,25	0	0,9	0
2	0,37	0,64	0	0,31
3	0	0,86	0,54	0
4	0,3	0,76	0	0

Tabla 1.8

a) Para el primer entrevistado:

$$C_1 = 1 - (1-0,25)^3 \cdot (1-0,9)^2 = 0,9958$$

$$G_1 = 3 \cdot 0,25 + 2 \cdot 0,9 = 2,55$$

$$\frac{C_1}{G_1} = 2,5608 \Rightarrow E_1 = 2 ; D_1 = 0,5608$$

Reciben 2 contactos: $(1 - 0,5608) \cdot 0,9958 \cdot 750 \approx 328$ individuos.

Reciben 3 contactos: $0,5608 \cdot 0,9958 \cdot 750 \approx 419$ individuos.

Reciben 0 contactos: $(1 - 0,9958) \cdot 750 \approx 3$ individuos.

b) Para el segundo entrevistado:

$$C_2 = 1 - (1 - 0,37)^3 \cdot (1 - 0,64)^4 \cdot (1 - 0,31)^3 = 0,9986$$

$$G_2 = 3 \cdot 0,37 + 4 \cdot 0,64 + 3 \cdot 0,31 = 4,6$$

$$\frac{C_2}{G_2} = 4,6063 \Rightarrow E_2 = 4 ; D_2 = 0,6063$$

Reciben 4 contactos: $(1 - 0,6063) \cdot 0,9986 \cdot 750 \approx 295$ individuos.

Reciben 5 contactos: $0,6063 \cdot 0,9986 \cdot 750 \approx 454$ individuos.

Reciben 0 contactos: $(1 - 0,9986) \cdot 750 \approx 1$ individuos.

c) Para el tercer entrevistado:

$$C_3 = 1 - (1 - 0,86)^4 \cdot (1 - 0,54)^2 = 0,9999$$

$$G_3 = 4 \cdot 0,86 + 2 \cdot 0,54 = 4,52$$

$$\frac{C_3}{G_3} = 4,5204 \Rightarrow E_3 = 4 ; D_3 = 0,5204$$

Reciben 4 contactos: $(1 - 0,5204) \cdot 0,9999 \cdot 750 \approx 360$ individuos.

Reciben 5 contactos: $0,5204 \cdot 0,9999 \cdot 750 \approx 390$ individuos.

Reciben 0 contactos: $(1 - 0,9999) \cdot 750 \approx 0$ individuos.

d) Para el cuarto entrevistado:

$$C_4 = 1 - (1 - 0,3)^3 \cdot (1 - 0,76)^4 = 0,9988$$

$$G_4 = 3 \cdot 0,3 + 4 \cdot 0,76 = 3,94$$

$$\frac{C_4}{G_4} = 3,9445 \Rightarrow E_4 = 3 ; D_4 = 0,9445$$

Reciben 3 contactos: $(1 - 0,9445) \cdot 0,9988 \cdot 750 \approx 41$ individuos.

Reciben 4 contactos: $0,9445 \cdot 0,9988 \cdot 750 \approx 708$ individuos.

Reciben 0 contactos: $(1 - 0,9988) \cdot 750 \approx 1$ individuo.

La distribución de contactos final es la suma de los resultados parciales, como aparece en la **Tabla 1.9**:

Nº de contactos	Individuos que reciben esos contactos
0	5 (3+1+0+1)
1	0
2	328
3	460 (419+41)
4	1.363 (295+360+708)
5	844 (454+390)
TOTAL	3.000 individuos

Tabla 1.9

La cobertura, que es el porcentaje de los que al menos contactan una vez, es el 99,83% (2.995/3.000). Poco después de la introducción del TOM en España, Pellicer³³ ejemplificó el modo de calcular la cobertura y la distribución de contactos.

4. Todas estas magnitudes se pueden ponderar por el coeficiente de memorización β o porcentaje de impactados que recuerdan el mensaje en un contacto. El alcance memorizado, la distribución de contactos memorizados, etc, se calculan de manera análoga, agregando las coberturas individuales *memorizadas* de cada entrevistado i :

$$C_i \text{ (memorizada después de } n \text{ contactos)} = C_i \text{ (inicial)} \cdot [1 - (1 - \beta)^n]$$

³³ PELLICER, José. "Panorámica actual". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Págs. 248-9.

Nótese que si el coeficiente β se aplica a nivel agregado, los resultados son ligeramente distintos, como sucedería si el alcance memorizado se calcula multiplicando el alcance por $[1-(1-\beta)^n]$. El programa hace la hipótesis de que β es constante en todos los contactos. Suso³⁴, a falta de estudios autóctonos, recoge los valores que β tomaba en Francia en 1982, pero la prudencia aconseja tomarlos sólo como indicativos:

Medio	Formato del mensaje	Valor medio de β para un impacto
Cine	30 segundos	70-75%
Televisión	30 segundos	15-20%
Revistas	página, 4 colores	10%
Diarios	600 mm.	10%
Radio	4x3	5%
Exterior		2%

³⁴ SUSO, Javier. "La función tiempo en la planificación". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Pág. 112.

TOM2/CX EVALUACIÓN DE PLANES CON DISTRIBUCIÓN TEMPORAL

El programa **TOM2C** permite evaluar el alcance y la distribución de contactos de una campaña, según la distribuyamos a lo largo del tiempo. Además de los inputs del TOM2, habrá que definir las semanas donde se pasarán los anuncios y las semanas de descanso. Una ola u oleada es un conjunto de semanas con inserciones, entre periodos de descanso. Las primeras salidas son iguales que el TOM2, para cada ola como si se tratase de un plan independiente; después evalúa semana a semana el plan en su conjunto, con las siguientes estimaciones:

- Cobertura memorizada hasta esa semana. El coeficiente de ponderación M se aplica a la cobertura según una distribución de contactos dada, para hallar la distribución de contactos memorizados; siendo \bar{c} el nº medio de contactos de la campaña:

$$M = 1 - (1 - \beta)^{\bar{c}}$$

Ésta y las posteriores fórmulas se deben, según Pistoni³⁵, a Armand Morgensztern. El coeficiente de memorización corregido β' es una variante de β que tiene en cuenta la duración de la campaña (o la ola) y su intensidad (o número de contactos); llamando da a la duración de las semanas con inserciones, tiene la expresión:

$$\beta' = \frac{\beta}{1 + 0,1 \cdot \left[\left(\frac{da}{\bar{c}} \right)^{1,5} - 1 \right]}$$

El coeficiente de memorización θ se aplica a los individuos que, no habiendo memorizado el mensaje durante la ola anterior, son contactados de nuevo por la

³⁵ PISTONI DURÁN, LUIS. "Programas actuales". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Pág. 259.

siguiente. Tiene un valor superior a β ya que, los individuos que han olvidado presentan una permeabilidad de rememorización más alta, que aquellos que todavía no han entrado en contacto con el mensaje; siendo dd la duración de las semanas de descanso:

$$\theta = \beta \left[\frac{dd}{dd + 2(\bar{e}+1)(\bar{e}+2)} \right]$$

- Cobertura media memorizada: suma de las coberturas memorizadas semanales dividida por el número de semanas transcurridas.

- Acumulación de coberturas memorizadas.

- Cobertura remanente o residual: es la cobertura memorizada al terminar un periodo de descanso o al finalizar la campaña. El coeficiente de memorización residual r decrece en función de la duración de dicho periodo:

$$r = 1 - (1 - M)^{\left[\frac{2\bar{e}}{2\bar{e} + dd(dd+1)} \right]}$$

El programa **TOM2CX** es una continuación del TOM2C, para evaluar campañas que estén divididas en más de doce periodos de actividad, por lo que es obligatorio haber utilizado previamente el TOM2C; los resultados antes obtenidos se incluyen como input, junto todas las olas y periodos de descanso que se quieren adicionar. Entonces evalúa la cobertura memorizada, la cobertura media memorizada, etc, para la campaña adicionada, acumulando los resultados con los del TOM2C. La mayor utilidad de estos programas es la ayuda para distribuir la campaña en el tiempo, porque se pueden comparar varias distribuciones alternativas, con la condición de que tengan el mismo β y el mismo número total de semanas.

TOM2M EVALUACIÓN MULTIMEDIA

Permite evaluar un plan dirigido a dos o tres medios, pero no en el sentido habitual porque el planificador puede indicar el medio al que pertenece cada soporte seleccionado. Es opcional el desglose según alguna de las variables. Se pueden asignar distintos coeficientes de memorización a los medios. Genera los mismos resultados que el TOM2, y adicionalmente el alcance, la cobertura total y la cobertura exclusiva de cada medio. También reproduce, en tablas de doble entrada, el cruce de las distribuciones de contactos de cada pareja de medios: las sumas marginales de filas y columnas son las distribuciones de contactos de cada medio.

TOM3 OPTIMIZACIÓN DE PLANES

Este programa construye el plan "óptimo" a partir de las restricciones impuestas por el planificador. Se basa en un procedimiento heurístico, como los empleados por los modelos de simulación Ad-Me-Sim, ADMOD y Calculus-Mediac (véase **IV Parte, §C**). Pretende maximizar: (1) la eficacia, en función de la cobertura y de la distribución de contactos según una curva de respuesta, y (2) el rendimiento del presupuesto.

El planificador indica los soportes sobre los que se busca la "mejor" combinación, y señala si son simplemente *candidatos*, o son *impuestos* porque se incluirán obligatoriamente en el mejor plan. Cada soporte, además de la habitual ponderación (*media-weight*), puede tener restricciones sobre el número máximo y mínimo de inserciones; el límite máximo puede depender, por ejemplo, por la duración de la campaña; el límite mínimo es, por defecto, una inserción. Por tanto, el soporte seleccionado incluye siempre un número de inserciones dentro del rango marcado entre el máximo y el mínimo. Además el planificador debe establecer el *paso de progresión* de cada soporte, es decir, el número de inserciones que aumentará ese soporte cada vez que sea elegido para el plan. Como el proceso de "optimización" puede ser largo, el programa tiene una opción para indicar desde donde se imprime la salida.

Bajo la denominación de *curva de respuesta*, el programa solicita al usuario la ponderación de cada nivel de la distribución de contactos (valor 0 ponderación mínima y valor 1 ponderación máxima).

Con la anterior información el programa puede construir el plan en sucesivas interacciones hasta gastar el presupuesto previsto. La primera interacción incluye los soportes *impuestos*; en cada sucesiva interacción se selecciona el soporte *candidato* que aporta mayor rendimiento marginal (incremento de cobertura entre incremento de presupuesto), tanto con las inserciones del *paso de progresión* como con el doble de esas inserciones, eligiendo la opción que aporte más. Una interacción supone entonces

calcular las coberturas (simulando individuo a individuo) del doble de todas las combinaciones de soportes candidatos: por ello el tiempo de solución se eleva cuando hay muchos candidatos, aunque el ordenador sea muy potente. Cada interacción se presenta con los siguientes resultados:

- Presupuesto consumido hasta ese momento.
- Alcance, cobertura, coste por mil impactos, coste por mil impactados, frecuencia, GRP's.
- Cobertura ponderada, tras aplicar el coeficiente de memorización o bien la ponderación de la curva de respuesta.
- Distribución de contactos.
- Eficacia del plan: índice para describir la proximidad al plan ideal. El plan "óptimo" tomaría el valor 100. La eficacia de un determinado plan se determina a partir de su cobertura (a) y de su porcentaje de cobertura ponderada sobre la cobertura (b), aplicando el teorema de Pitágoras, como se observa en la **Figura 1.6**.

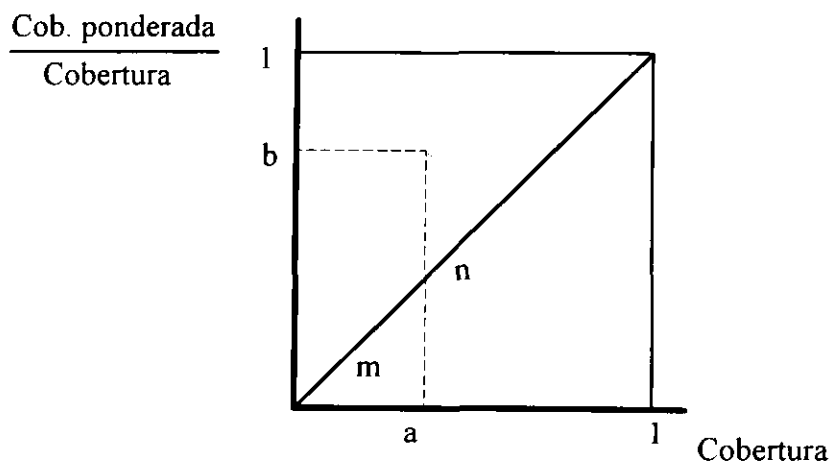


Figura 1.6

Gráficamente, la eficacia se define como la proximidad entre el punto proyectado sobre la diagonal y el punto [1,1]. La eficacia del plan aumenta según nos desplazamos hacia el ángulo superior derecha, es decir, cuando crece la cobertura y la cobertura

ponderada supone un mayor porcentaje de esa cobertura. Algebraicamente toma esta expresión:

$$\text{Eficacia} = \frac{m+n}{\sqrt{2}} = \frac{a\sqrt{2} + \frac{b-a}{\sqrt{2}}}{\sqrt{2}} = \frac{a+b}{2} = \frac{\text{Cobertura} + \frac{\text{Cob. ponderada}}{\text{Cobertura}}}{2}$$

- Rendimiento marginal: cociente entre el incremento de eficacia y el incremento de presupuesto. Si es mayor que 0, indica que el plan se puede mejorar, invirtiendo mayor parte del presupuesto. Cuando se acerca mucho al valor 0, altos incrementos en el presupuesto sólo consiguen bajos incrementos en la eficacia. El rendimiento marginal podría ser negativo cuando perdiera eficacia, en el caso de una distribución de contactos ponderada.

TOM4 GENERACIÓN DE PLANES ÓPTIMOS

Una vez determinado el plan "óptimo", el TOM4 genera una serie de planes alternativos a ese. Los inputs son iguales que en el TOM3. El proceso es igual que antes, hasta que se consume el 98% del presupuesto, momento en el que empieza la generación de alternativas: se elimina una inserción y se coloca otra, rondando el límite presupuestario. Ahora no se recoge la eficacia y el rendimiento marginal.

TOM5 CONSULTAS A FICHERO: TABULACIONES

Es un programa para analizar cualquiera de las variables disponibles en el fichero, bien de una en una, o bien cruzando varias. Hay que designar la/-s *variable/-s de ventilación*, cuyas alternativas se van a cruzar con la variable objeto de estudio; la salida presenta en filas las variables de ventilación y en la cabecera la variable explicada, organizada en los llamados *grupos estudiados*.

2. Programas disponibles únicamente en Mainframe

Otros programas TOM todavía no han sido adaptados al ordenador personal en la versión TOM-Micro. Según nos han comentado, no está prevista su adaptación porque no se utilizan mucho. No obstante los citaré brevemente porque algunos ayudan a comprender el funcionamiento del modelo TOM y para completar el estudio del paquete informático.

TOMCOB ANÁLISIS DE COBERTURAS INSERCIONES

Como información inicial, hay que definir el fichero de uso, el público objetivo, el soporte y las n inserciones a analizar. El programa, como siempre, devuelve primero una estimación del público objetivo (real y ponderado) y el número de entrevistados incluidos en ese *target*. Después, específicamente, el TOMCOB estima:

- las coberturas acumuladas y las distribuciones de contactos generadas por 1, 2,..., n inserciones. Obviamente suponiendo una inserción por ejemplar del soporte. A diferencia de anteriores programas, éste no simula el comportamiento individual, sino que trabaja con información a nivel agregado. Los entrevistados están agrupados con las probabilidades que resultan del cociente entre el hábito y la audiencia del último periodo (**III Parte, §C.5**): a cada grupo se le aplica el modelo binomial (**III Parte, §B.1**), y finalmente se suman las distribuciones de contactos de cada grupo.

- la cobertura máxima (real y ponderada) del soporte, que resulta de dividir los que han contactado alguna vez con el soporte entre la población objetiva (real y ponderada, respectivamente).

- porcentaje de la cobertura máxima que se va cubriendo con la cobertura acumulada de 1, 2,..., n inserciones.

- OTS (frecuencia) con 1, 2,..., n inserciones: media de contactos recibidos por los individuos alcanzados.

TOM4 GENERACIÓN DE PLANES ÓPTIMOS

Una vez determinado el plan "óptimo", el TOM4 genera una serie de planes alternativos a ese. Los inputs son iguales que en el TOM3. El proceso es igual que antes, hasta que se consume el 98% del presupuesto, momento en el que empieza la generación de alternativas: se elimina una inserción y se coloca otra, rondando el límite presupuestario. Ahora no se recoge la eficacia y el rendimiento marginal.

TOM5 CONSULTAS A FICHERO: TABULACIONES

Es un programa para analizar cualquiera de las variables disponibles en el fichero, bien de una en una, o bien cruzando varias. Hay que designar la/-s *variable/-s de ventilación*, cuyas alternativas se van a cruzar con la variable objeto de estudio; la salida presenta en filas las variables de ventilación y en la cabecera la variable explicada, organizada en los llamados *grupos estudiados*.

2. Programas disponibles únicamente en Mainframe

Otros programas TOM todavía no han sido adaptados al ordenador personal en la versión TOM-Micro. Según nos han comentado, no está prevista su adaptación porque no se utilizan mucho. No obstante los citaré brevemente porque algunos ayudan a comprender el funcionamiento del modelo TOM y para completar el estudio del paquete informático.

TOMCOB ANÁLISIS DE COBERTURAS INSERCIONES

Como información inicial, hay que definir el fichero de uso, el público objetivo, el soporte y las n inserciones a analizar. El programa, como siempre, devuelve primero una estimación del público objetivo (real y ponderado) y el número de entrevistados incluidos en ese *target*. Después, específicamente, el TOMCOB estima:

- las coberturas acumuladas y las distribuciones de contactos generadas por 1, 2,..., n inserciones. Obviamente suponiendo una inserción por ejemplar del soporte. A diferencia de anteriores programas, éste no simula el comportamiento individual, sino que trabaja con información a nivel agregado. Los entrevistados están agrupados con las probabilidades que resultan del cociente entre el hábito y la audiencia del último periodo (**III Parte, §C.5**): a cada grupo se le aplica el modelo binomial (**III Parte, §B.1**), y finalmente se suman las distribuciones de contactos de cada grupo.

- la cobertura máxima (real y ponderada) del soporte, que resulta de dividir los que han contactado alguna vez con el soporte entre la población objetiva (real y ponderada, respectivamente).

- porcentaje de la cobertura máxima que se va cubriendo con la cobertura acumulada de 1, 2,..., n inserciones.

- OTS (frecuencia) con 1, 2,..., n inserciones: media de contactos recibidos por los individuos alcanzados.

Como señala Zambrano³⁶, dependiendo de la distribución de probabilidades, "podemos obtener unas distribuciones de contacto y unos porcentajes sobre la cobertura máxima muy diferentes"; entonces, este programa, a partir de la estructura de audiencia del soporte, puede ayudar a determinar el número adecuado de inserciones, con la presión suficiente y, desde la perspectiva contraria, sin la saturación no rentable.

TOMSEG SEGMENTACIÓN

El TOMSEG es una programa de segmentación que utiliza la técnica A.I.D. (Automatic Interaction Detection³⁷). Va dividiendo la población en segmentos, en función de unas variables independientes que explican la variable dependiente. Entre las variables independientes se elige aquella que hace subgrupos más heterogéneos entre sí y más homogéneos consigo mismo.

Después de introducir los inputs habituales, hay que definir la variable dependiente (la audiencia de un soporte, el consumo de un producto, etc.) y las variables independientes (cualquiera de la variables de clasificación, consumo, equipamiento, etc). Se debe indicar si la variable es ordinal, para no segmentar aleatoriamente cuando existe una secuencia de orden relacionada (directa o inversamente) con la variable dependiente: en el caso de la lectura de diarios, serían ordinales la clase social y el nivel de estudios, pero no el sexo, la edad o el hábitat. Permite recodificar, agrupando las alternativas de las variables: por ejemplo, la edad tiene siete ítems (14-19; 20-24; 25-29; ... ; 60-64; >64) que se pueden agrupar en tres (14-34; 35-54; >54). Las salidas ofrecen las siguientes magnitudes:

³⁶ ZAMBRANO, LOURDES. "La probabilización en el Estudio General de Medios". Ponencia presentada en el 2º Seminario sobre Medios Impresos (54º Seminario de AEDEMO). Madrid. Diciembre 1990. Resumen de ponencias. Pág. 53.

³⁷ Véase: SONQUIST, J.A. "Multivariate model building". Survey Research Center. Ann Arbor. Michigan. 1970.

- Media y desviación típica de la variable explicada (en función de la escala utilizada en la codificación).

- Descripción de las sucesivas divisiones: la población objetivo se divide en dos segmentos; cada uno de ellos se divide en otros dos; y así sucesivamente, pero siempre cada grupo se divide en *dos* subgrupos. En cada partición, el programa encuentra la *variable de ruptura*, aquella que divide en los dos grupos con mayor diferencia de medias; tiene en cuenta tanto la variable como la alternativa por donde cortar, evaluando el corte óptimo entre todas las combinaciones posibles.

- Los segmentos finales se presentan ordenados de mayor a menor valor medio en la variable explicativa, junto con la siguiente información: descripción (p.ej.: hombres; 14-34 años,...), efectivos (nº de personas que incluye y tanto por mil sobre el público objetivo), media de la variable explicativa, efectivos acumulados, media acumulada de la variable explicativa y tanto por mil de la marca, que en el último segmento alcanzará la unidad:

$$\% \text{ MARCA} = \frac{\text{media acumulada segmento} \times \text{población acumulada segmento}}{\text{media total} \times \text{población total}}$$

Finalmente, bajo el nombre de arborescencias, se representan las sucesivas particiones en un diagrama de árbol; facilitan una visión global y completa del conjunto aunque, en mi opinión, la presentación de esta salida debería ser más clara.

OTROS PROGRAMAS TOM

Incluimos en este epígrafe varios programas que merecen menor importancia:

- TOMPER (Perfiles sociodemográficos): ayuda en el tratamiento de los públicos objetivos, dando la distribución del *target* por variables sociodemográficas.

- TOMEST (Estructura de Medios): una vez definido el público objetivo, calcula la cobertura alcanzada por cada medio y las duplicaciones existentes entre los diversos medios.

- TOMFAC (Análisis Factorial de Correspondencias): sirve para mostrar las afinidades existentes entre dos conjuntos de variables. Se puede analizar el posicionamiento de una serie de soportes en función de las variables indicadas o, de la misma forma, el posicionamiento de un producto. Los resultados se incluyen en un gráfico formado por los dos factores que explican más varianza.

D) CONTRASTE FRENTE A UN PANEL

En anteriores apartados se indican los *outputs* generados por los programas TOM, como una valiosa información para las decisiones sobre medios. Muchos *outputs* se calculan exactamente a partir de las bases de datos (GRP's, alcance bruto, coste por mil impactos, etc) pero otros, dada la información inicial, sólo pueden ser estimados (alcance y distribución de contactos). La estimación, a diferencia del cálculo exacto, está sujeta a errores. La evaluación de tales errores es el objetivo de las siguientes páginas. No se trata simplemente de testar algunos resultados importantes del TOM, sino de chequear el fundamento mismo de la planificación de medios mediante estos programas.

Como señala Marc³⁸, en todo modelo de simulación se deberían "chequear un conjunto de campañas diversas, para encontrar si las distribuciones de contactos simuladas se ajustan a las distribuciones de contactos generadas en un panel". El panel recoge la exposición a los medios continuamente, de forma que se dispone de la evolución de la audiencia instante a instante. Se distingue así de los estudios periódicos, que en un momento miden el comportamiento de la audiencia durante todo un periodo. En un panel no hace falta estimar el alcance y la distribución de contactos, porque se pueden calcular exactamente. Por eso las estimaciones del TOM sobre un estudio periódico se pueden contrastar con las observaciones directas de un panel, del que previamente se extrajo el estudio periódico: así se evalúan los errores de estimación.

Desde hace varios años, en España funcionan los audímetros para Televisión y, exclusivamente en Cataluña, existe un Diario de Escucha de Radio. Muchos factores influyen para que en el futuro existan paneles de Diarios, Revistas,...: [1] la financiación exigida debe corresponder a un volumen publicitario suficiente (véase §A); [2] los altos costes de esos estudios deben ser rentabilizados; [3] todas las partes deben ponerse de acuerdo; etc. Sin pretender analizar tales cuestiones, sí se mide la insuficiencia de la

³⁸ MARC, M.: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Pág. 16.

investigación periódica: se puede evaluar entonces la conveniencia de los paneles de audiencia, a partir de la rentabilidad esperada por una mejor planificación.

1. Descripción del experimento e hipótesis implícitas

Se ha generado un *panel de audiencia*, sobre una muestra de 1.000 individuos, que recoge la exposición a 20 soportes, durante 18 periodos (semanas, días, franjas horarias,...). Cuando el individuo contacta con el soporte en un periodo, se nota como 1, y su contrario como 0. La información codificada se organiza en 18 hojas, una por periodo, representándose los individuos en las filas ($I_1, I_2, I_3, \dots, I_{1000}$) y los soportes en las columnas ($S_1, S_2, S_3, \dots, S_{20}$). La **Figura 1.7** lo muestra gráficamente.

Cuando la investigación de medios se ha de limitar a estudios periódicos, como el EGM, podemos utilizar diversos algoritmos para estimar el alcance y la distribución de contactos. Hemos comparado entonces los errores del TOM con los de Metherringham (modelo de fórmula) y del MARKSEL (modelo de simulación).

La llamada *ola* correspondería a un estudio periódico, y contiene las probabilidades medias de exposición para cada individuo (en filas) y cada soporte (en columnas). Se extrae del panel dividiendo, en cada caso, el número de contactos efectuados entre las 18 exposiciones posibles: este método, que coincide con el *actuarial*, es el apropiado medir la probabilidad a partir de un panel, igual que el *método del último periodo* es el que mejor se adapta al estudio periódico del EGM (**III Parte, § C.3 y C.5**). Téngase en cuenta que se están obviando los fallos de memoria y de interpretación de los entrevistados, la audiencia paralela y replicada, etc.

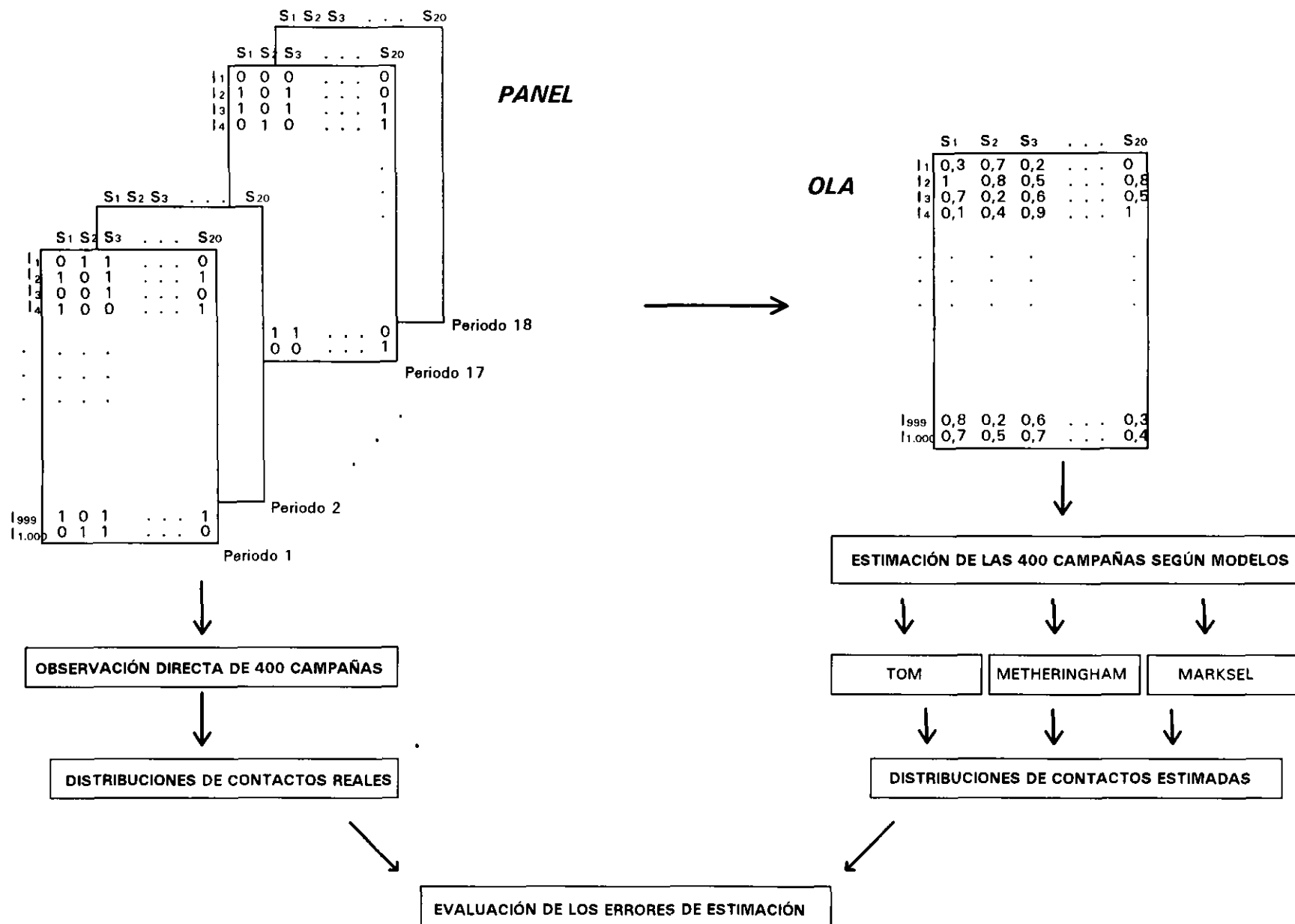


Figura 1.7

La exposición se ha considerado un fenómeno aleatorio: en el *panel*, los valores 1 (exposición) y 0 (no exposición) se han determinado a través de la función aleatoria del programa Excel 3.0. Sin hacer ninguna restricción, todas las probabilidades medias de la *ola* estarían cerca de 0'5: por eso hemos dividido los individuos en 9 grupos iguales, que suponemos con distintos niveles de audiencia: sus probabilidades medias rondarían respectivamente 0'1, 0'2, ... hasta 0'9. Por ejemplo, en los 110 individuos del quinto grupo, la fórmula³⁹ empleada ha sido =ENTERO(ALEATORIO()*2), con lo que se obtiene aproximadamente mitad de valores 1 y mitad 0 (valor medio más probable igual a 0'5); en el octavo grupo, para los 110 individuos se utiliza la fórmula =1-ENTERO(ALEATORIO()/0,8), obteniéndose aproximadamente un 80% de valores 1 y un 20% de valores 0 (valor medio más probable de 0'8).

Calculadas las probabilidades individuales de la *ola*, se les aplica una a una los algoritmos de los modelos TOM, Metheringham y MARKSEL, para estimar la distribución contactos de una combinación de soportes, correspondiente a una campaña. Se pretende comparar las distribuciones estimadas con la realidad directamente observada en el panel, midiendo la dimensión y tendencia de los errores. Lo mismo se hace con el alcance. Se evalúan los resultados de 400 campañas, construidas semialeatoriamente: se les ha impuesto, por grupos de 20, unas características comunes en cuanto a la duración de la campaña, el número de soportes utilizados y las inserciones resultantes (véase la **Tabla 1.10**); se supone una sola inserción por soporte y periodo; los anuncios aparecen ininterrumpidamente mientras dura la campaña; el inicio de la campaña se determina en cada caso aleatoriamente, con la restricción de que todas las inserciones se coloquen en los 18 periodos que ocupa la *ola*; por tanto, en cada soporte seleccionado se incluyen exactamente tantas inserciones como periodos dura la campaña; los soportes, en cada caso, se seleccionan aleatoriamente entre los 20 posibles.

³⁹ La descripción de las funciones habitualmente aparece en todos los manuales del programa, por ejemplo: BÖHMER, Martin. "EXCEL 3. Acceso rápido". DATA BECKER. Editado en castellano por Marcombo Boixareu Editores. Barcelona. 1991.

Nº de campañas según cada tipo	Duración	Nº de soportes de la campaña	Nº de inserciones de la campaña
20 casos	2 periodos	2	4
20 casos	2 periodos	4	8
20 casos	2 periodos	6	12
20 casos	2 periodos	8	16
20 casos	2 periodos	10	20
20 casos	3 periodos	2	6
20 casos	3 periodos	3	9
20 casos	3 periodos	4	12
20 casos	3 periodos	5	15
20 casos	3 periodos	6	18
20 casos	4 periodos	2	8
20 casos	4 periodos	3	12
20 casos	4 periodos	4	16
20 casos	4 periodos	5	20
20 casos	5 periodos	2	10
20 casos	5 periodos	3	15
20 casos	5 periodos	4	20
20 casos	6 periodos	2	12
20 casos	6 periodos	3	18
20 casos	7 periodos	2	14

Tabla 1.10: Tipología de las campañas.

El algoritmo utilizado por el TOM ya ha sido expuesto y ejemplificado anteriormente (§C.1). Después se hace lo mismo con el modelo de Metherringham (II Parte, §A.5). Ahora vamos a explicar otro de los algoritmos empleados en este experimento, "el que utiliza la versión del MARKSEL actualmente en uso en España"⁴⁰. Debe clasificarse dentro de los modelos de simulación (cfr. IV Parte, §B) ya que estima el alcance y la distribución de contactos, individuo a individuo, agregando luego los resultados. Se basa en la hipótesis fundamental de la homogeneidad de soportes, convirtiendo en una probabilidad media la probabilidad de exposición a varios soportes, como si se redujera a un hipotético soporte medio. Siendo p_{ij} la probabilidad de que el individuo i contacte con el soporte j , y r_j el nº de inserciones del soporte j , la probabilidad media \bar{p}_i se calcula:

⁴⁰ LAMAS ALONSO, Carlos: "Eso de las evaluaciones no lo veo claro". *Investigación y Marketing*. Asociación Española de Estudios de Mercado, Marketing y Opinión (AEDEMO). Nº 42. Julio 1993. Pág. 61.

$$\bar{p}_i = \frac{\sum_{j=1}^m r'_{ij} \cdot p_{0j}}{\sum_{j=1}^m r'_{ij}} \quad \begin{array}{ll} r'_{ij} = r_{ij} & \text{si } p_{0j} \neq 0 \\ r'_{ij} = 0 & \text{si } p_{0j} = 0 \end{array}$$

Por hipótesis, la exposición al soporte medio en un momento determinado es independiente de la exposición en otro momento. La distribución de contactos se determina a partir de una distribución binomial: siendo N el número total de inserciones de la campaña, la probabilidad de que el individuo i esté expuesto exactamente a k inserciones vale:

$$P(x = k / N, \bar{p}_i) = \binom{N}{k} \cdot \bar{p}_i^k \cdot (1 - \bar{p}_i)^{N-k}$$

Entonces, tal individuo tiene esta probabilidad de contactar al menos con una de las inserciones (probabilidad de ser alcanzado por la campaña):

$$1 - P(x = 0 / N, \bar{p}_i) = 1 - (1 - \bar{p}_i)^N$$

A continuación se incluye un **ejemplo** de este algoritmo, sobre el mismo enunciado utilizado en el TOM (**Tabla 1.8**). Primero calculamos, para los cuatro individuos, las probabilidades de exposición al hipotético soporte medio:

$$\bar{p}_1 = \frac{3 \cdot 0'25 + 2 \cdot 0'9}{3 + 2} = 0'51 \quad \bar{p}_2 = \frac{3 \cdot 0'37 + 4 \cdot 0'64 + 3 \cdot 0'31}{3 + 4 + 3} = 0,46$$

$$\bar{p}_3 = \frac{4 \cdot 0'86 + 2 \cdot 0'54}{4 + 2} \approx 0,75 \quad \bar{p}_4 = \frac{3 \cdot 0'3 + 4 \cdot 0'76}{3 + 4} \approx 0,56$$

Las 750 personas representadas por cada individuo recibirán tantos contactos como resulte de aplicar la binomial: $750 \cdot \binom{12}{k} \cdot \bar{p}_i^k \cdot (1 - \bar{p}_i)^{12-k}$. La distribución de

contactos de la población total es la suma de las distribuciones de contactos individuales, como aparece en la **Tabla 1.11**.

Contactos (k)	Individuo 1 ($\bar{p}_1=0'51$)	Individuo 2 ($\bar{p}_2=0'46$)	Individuo 3 ($\bar{p}_3=0'75$)	Individuo 4 ($\bar{p}_4=0'56$)	DISTRIBUCIÓN TOTAL
0	0	0	0	0	1
1	2	5	0	1	7
2	10	22	0	4	36
3	36	63	0	17	116
4	83	120	2	50	255
5	139	164	8	102	413
6	169	163	29	154	514
7	151	119	75	170	514
8	98	63	143	137	440
9	45	24	194	78	341
10	14	6	177	30	228
11	3	1	98	7	109
12	0	0	25	1	26
Población representada	750	750	750	750	3000

Tabla 1.11: Ejemplo del algoritmo utilizado por el programa MARKSEL.

Antes de mostrar los resultados, haré algunas consideraciones metodológicas: me he cuestionado la **validez** de los resultados, ya que la aleatoriedad de las exposiciones puede tergiverar el habitual funcionamiento de los algoritmos. En concreto, nuestro panel no tiene en cuenta: [1] la correlación entre *soportes* que, por ejemplo, se da entre las audiencias de dos programas de TV con contenidos similares; [2] la correlación entre *periodos* que se manifiesta, entre otros casos, en la fidelidad de los lectores habituales de un periódico o revista; [3] las correlaciones entre *individuos* que supone, por ejemplo, que dos individuos de características similares tengan preferencia por un tipo de programas de radio; etc. Al mismo tiempo me he preguntado si los modelos tienen en cuenta tales correlaciones, como se dan en la realidad. Vamos a analizar entonces las hipótesis implícitas, en los modelos, en el panel aquí experimentado y en la realidad; ahora se citan pero luego se comentan ampliamente, al inicio de cada Parte.

HIPÓTESIS	Estacionariedad	Población	Soportes	Duplicación	Acumulación
Modelo TOM	Estacionario	Heterogénea	Heterogéneos	Aleatoria	Aleatoria
Modelo de Metherringham	Estacionario	Heterogénea	Homogéneos	Beta Media	Beta Media
Modelo MARKSEL	Estacionario	Heterogénea	Homogéneos	Aleatoria	Aleatoria
Panel experimentado	No Estacionario	Heterogénea	Heterogéneos	Aleatoria	Aleatoria
Realidad	No Estacionario	Heterogénea	Heterogéneos	No Aleatoria	No Aleatoria

La exposición es una realidad compleja; los modelos la tratan como un fenómeno probabilístico y hacen hipótesis que simplifican la realidad: así la duplicación aleatoria supone asumir que no existe relación entre las exposiciones a uno y otro soportes. Es válido contrastar los modelos con este panel, porque se hacen hipótesis similares en cuanto las exposiciones. Mostrada la viabilidad del experimento, hay que señalar su conveniencia: los modelos son comparables, porque sus algoritmos deben estimar el alcance y la distribución de contactos, ajustándose a los datos originales; dicho de otra forma, la aleatoriedad de la exposición no exonera al algoritmo de una buena estimación; al contrario, la estimación debe ser habitualmente mejor, porque los datos se basan en hipótesis semejantes.

Este contraste permite la *verificación interna* de los algoritmos, mientras que la contrastación con un panel real aportaría la *verificación externa* (**Introducción**). El primero aporta conclusiones más generales porque analiza la estructura interna, mientras que las conclusiones del segundo son generalizables en función de la semejanza con los datos contrastados. Supongamos que se hubiera contrastado con alguno de los paneles disponibles de TV o Radio; sea cual fuere su diseño, no se pueden recoger todos los comportamientos posibles; siempre se podría encontrar otro panel con nuevos comportamientos, que darían lugar a distintas distribuciones de contactos; por eso, una buena forma de generalizar los resultados buscados con el experimento, es que el fenómeno de la exposición sea considerado aleatorio. Posteriores investigaciones

deberían contrastar con paneles reales, y medir la precisión de los algoritmos en cada medio y en las diversas circunstancias que se plantean en la planificación de medios.

Los modelos TOM, Metherringham y MARKSEL se basan en la hipótesis de la estacionariedad de la audiencia, es decir, la probabilidad de exposición se mantiene constante durante el tiempo. Disponiendo de un panel no habría que hacer esta hipótesis, porque se conoce directamente la evolución de la audiencia. Algunos autores intentaron medir el efecto de la estacionariedad: Schreiber⁴¹ llegó a la conclusión de que muchos modelos subestimaban el alcance y sobrestimaban la frecuencia porque ignoraban la inestabilidad de los hábitos de exposición; Chandon⁴² verificó esa conclusión respecto al alcance, pero no encontró una tendencia definida en la distribución de contactos; también observó que si la exposición se considera aleatoria, la subestimación del alcance será menor. Los resultados de este experimento, que se exponen en siguientes apartados, son también un test de la hipótesis de estacionariedad.

Las campañas observadas en este panel tienen distribuciones en forma de campana, debido al Teorema Central del Límite. Como comentaremos luego, las distribuciones de contactos pueden tomar formas diversísimas, y el modelo debería adaptarse a todos los casos.

Con la ayuda de la Prof. D^a Pilar Martín-Guzmán, he descubierto que el experimento adolece de una restricción: al agregar las probabilidades de exposición, como hace Metherringham, todos los soportes tienen probabilidades de exposición similares, cercanas a 0'5; igualmente todas las duplicaciones *entre*-soportes e *intra*-soportes tienen unas probabilidades cercanas a 0'31. El caso más general podía plantear soportes con distintas audiencias a nivel agregado, como se habían distinguido varios niveles de audiencia entre los individuos. Hay que tener en cuenta que las conclusiones,

⁴¹ SCHREIBER, Robert J. "Instability in Media Exposure Habits". *Journal of Advertising Research*. Vol. 14. N° 2. Abril 1974. Págs. 13-7.

⁴² CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 288-92.

sobre todo en cuanto al modelo de Metherringham, han sido deducidas a partir de soportes con tales características.

2. Precisión del alcance estimado

Como ya se definió, el alcance es la suma de los individuos expuestos a una determinada campaña; se relaciona inmediatamente con la distribución de contactos, porque es la suma de los que reciben al menos un contacto o, por exclusión, todos los que no reciben cero contactos. Hemos obtenido, para las 400 campañas, el alcance observado directamente en el panel y el alcance estimado por los modelos TOM, Metheringham y MARKSEL. Para evaluar la precisión de las estimaciones, se ha optado por la siguiente medida del error:

$$\text{error de estimación (\%)} = \text{alcance estimado (\%)} - \text{alcance observado (\%)}$$

Esta diferencia también se toma en valor absoluto, para estudiar tanto el signo como la dimensión del error. Todas las magnitudes están en porcentaje sobre el público objetivo: el alcance, así considerado, se conoce como cobertura. Evidentemente, estas medidas se aplican igual a todos los modelos.

Como muestran las siguientes tablas, el TOM estima mejor en conjunto el alcance que el MARKSEL, y éste a su vez mejor que Metheringham. Comparativamente, no existen importantes diferencias entre los modelos. Esta tendencia se observa tanto en valores simples como en absolutos. Los errores son casi exclusivamente de sobrestimación (muy poca diferencia entre los valores simples y absolutos); sólo el modelo de Metheringham, para pocas inserciones, presenta desviaciones positivas y negativas, que anulan el valor medio.

En las sucesivas tablas se recogen los resultados de forma resumida: las medias de los errores según los grupos característicos de las campañas; verticalmente se organiza en tres grandes columnas, con los resultados medios de cada modelo; horizontalmente están los factores que definen las campañas. En la última fila aparecen siempre los errores medios para las 400 casos.

Nº de inserciones	T O M		Metherringham		MARKSEL		Nº de casos
	Simples	Absolutos	Simples	Absolutos	Simples	Absolutos	
4	-0,54%	0,78%	0,37%	0,76%	-0,82%	0,92%	20
6	-0,84%	0,97%	-0,07%	0,75%	-1,02%	1,11%	20
8	-0,66%	0,84%	-0,31%	0,69%	-0,82%	0,93%	40
9	-0,68%	0,88%	-0,56%	0,85%	-0,83%	0,99%	20
10	-1,00%	1,00%	-0,65%	0,72%	-1,09%	1,09%	20
12	-0,62%	0,69%	-0,75%	0,80%	-0,72%	0,77%	80
14	-0,99%	0,99%	-0,82%	0,83%	-1,04%	1,04%	20
15	-0,57%	0,61%	-1,02%	1,04%	-0,65%	0,68%	40
16	-0,26%	0,35%	-0,89%	0,89%	-0,34%	0,40%	40
18	-0,43%	0,52%	-0,99%	0,99%	-0,49%	0,56%	40
20	-0,32%	0,42%	-1,07%	1,08%	-0,38%	0,45%	60
TODAS	-0,57%	0,66%	-0,72%	0,88%	-0,67%	0,74%	400

Tabla 1.12: Errores medios en los alcances estimados por TOM, Metherringham y MARKSEL (según el nº de inserciones).

La **Tabla 1.12** muestra que las inserciones influyen en el tamaño de los errores de todos los métodos, aunque no se observa una tendencia definible: no existe una relación clara entre el nº de inserciones y la evolución de los errores; sólo se ve que, a partir de la decimoquinta inserción, los errores disminuyen significativamente con el TOM y el MARKSEL. Lo habitual es que los valores simples estén próximos a los absolutos, para cualquier inserción, indicando la tendencia general a la subestimación; pero no sucede así en el modelo de Metherringham, en niveles inferiores a seis inserciones.

Nº de periodos	T O M		Metherringham		MARKSEL		Nº de casos
	Simples	Absolutos	Simples	Absolutos	Simples	Absolutos	
2	-0,30%	0,49%	-0,55%	0,85%	-0,45%	0,57%	100
3	-0,51%	0,63%	-0,66%	0,87%	-0,63%	0,72%	100
4	-0,56%	0,64%	-0,71%	0,85%	-0,65%	0,70%	80
5	-0,78%	0,80%	-0,96%	0,99%	-0,85%	0,87%	60
6	-0,85%	0,85%	-0,88%	0,91%	-0,91%	0,91%	40
7	-0,99%	0,99%	-0,82%	0,83%	-1,04%	1,04%	20
TODAS	-0,57%	0,66%	-0,72%	0,88%	-0,67%	0,74%	400

Tabla 1.13: Errores medios en los alcances estimados por TOM, Metherringham y MARKSEL (según el nº de periodos).

Habíamos comentado que el nº de inserciones es el producto de los periodos por los soportes. Veamos si estos factores explican las divergencias encontradas en las inserciones. La **Tabla 1.13**, con la misma estructura de la anterior, agrupa las campañas por filas según el nº de periodos. Es significativo que los errores aumenten con los periodos, conforme se descende en la tabla. En valor absoluto, los errores medios del TOM y MARKSEL se duplican progresivamente entre 2 y 7 periodos, pero los de Metheringham permanecen estables; sin embargo, en valores simples, aumentan todos pero en mayor medida el TOM y el MARKSEL. El análisis de los datos también enseña que la mayor parte de los casos de sobrestimación se dan entre 2 y 3 periodos.

Nº de soportes	T O M		Metheringham		MARKSEL		Nº de casos
	Simple	Absolutos	Simple	Absolutos	Simple	Absolutos	
2	0,89%	0,97%	0,36%	0,77%	1,02%	1,06%	120
3	0,67%	0,75%	0,83%	0,92%	0,77%	0,83%	80
4	0,47%	0,56%	0,79%	0,88%	0,58%	0,63%	80
5	0,34%	0,40%	1,02%	1,02%	0,42%	0,45%	40
6	0,28%	0,46%	0,92%	0,95%	0,38%	0,53%	40
8	0,10%	0,21%	0,88%	0,88%	0,19%	0,25%	20
10	0,07%	0,27%	0,99%	1,02%	0,14%	0,30%	20
TODOS	0,57%	0,66%	0,72%	0,88%	0,67%	0,74%	400

Tabla 1.14: Errores medios en los alcances estimados por TOM, Metheringham y MARKSEL (según el nº de soportes).

El nº de soportes también marca la tendencia errática de los alcances estimados: TOM y MARKSEL mejoran sistemáticamente al incrementarse los soportes, mientras que en Metheringham la mejora es suave y más difusa (véase **Tabla 1.14**). En conjunto, la influencia del factor *soportes* es mayor que la del factor *periodos*; como tienen tendencias contrarias, el producto (nº de inserciones) manifiesta diferencias en sus alternativas, pero sin una tendencia definida. Este comportamiento del TOM avala la contrastación de modelos realizada en la **II Parte**, con muchas revistas y en un sólo periodo de tiempo: así es como mejor estima el alcance, y además siempre lo hace mejor que los demás modelos.

3. Precisión en la estimación de la distribución de contactos

La distribución de contactos es un valioso instrumento de la planificación de medios, que distribuye una población determinada según el n° de exposiciones recibidas en una campaña. Cada contacto (0, 1, 2, ..., hasta las n inserciones de la campaña) tiene asociado una frecuencia o porcentaje de la población: por eso también se llama distribución de frecuencias. Obviamente, la suma de las frecuencias debe ser la unidad. Se han calculado exactamente las distribuciones de contactos generadas por las 400 campañas en el panel. Luego se han comparado con las 400 distribuciones de contactos estimadas por el TOM, Metheringham y el MARKSEL. Los errores de cada modelo se han medido de una forma sencilla y rigurosa, que permite comparar campañas con un n° heterogéneo de inserciones.

<u>Contactos</u>	<u>Frecuencia estimada</u>	<u>Frecuencia observada</u>	<u>Error absoluto por cont.</u>
0	f'_0	f_0	
1	f'_1	f_1	$ f'_1 - f_1 $
2	f'_2	f_2	$ f'_2 - f_2 $
...
n	f'_n	f_n	$ f'_n - f_n $
<hr/>			
	$\sum_{i=0}^n f'_i = 1$	$\sum_{i=0}^n f_i = 1$	$0 \leq \sum_{i=1}^n f'_i - f_i \leq 1$

El error absoluto total de una campaña, $\sum_{i=1}^n |f'_i - f_i|$, se mide también en porcentaje sobre el público objetivo. Así llegamos a 400 valores, tantos como campañas, en cada uno de los tres modelos. Esta información va a ser la base de los siguientes análisis.

Resulta muy significativo que la suma total de errores del TOM se reduzca casi a la mitad con Metheringham y casi al 40% con MARKSEL. En la práctica corriente, a muchos les resultaría difícil aceptar que un modelo de fórmula pueda duplicar la

precisión del modelo de simulación TOM; parecería más verosímil que otro modelo de simulación, con un algoritmo perfeccionado, mejorara así los resultados; sin embargo, los datos inequívocos de esta investigación colocan al modelo de Metheringham muy por delante del TOM; no en vano, el modelo de Metheringham y otros directamente relacionados, son utilizados frecuentemente por los planificadores⁴³. Es necesario, entonces, que en nuestro país se continúen desarrollando las investigaciones, y su difusión contribuya a una planificación más eficiente, en todos los niveles, tal como nos proponíamos en la **Introducción**.

Igual que con el alcance, hemos analizado la relación entre los errores de estimación y las características de la campañas: nº de inserciones, nº de periodos y nº de soportes. En las tres siguientes tablas, se recoge la media y la desviación típica de los errores para cada grupo alternativo. La **Tabla 1.15** muestra la influencia variable del número de inserciones sobre los errores: esta relación es inversa en el TOM, pero directa en Metheringham y MARKSEL; en estos últimos, los errores aumentan sistemáticamente con las inserciones mientras que, en el TOM, primero disminuyen y luego aumentan desde la decimoquinta inserción.



⁴³ LECKENBY, John D. y KIM, Heejin. "How Media Directors View Reach/Frequency Estimation: Now and Decade Ago". *Journal of Advertising Research*. Vol. 34. Nº 5. Septiembre-Octubre de 1994. Pág. 13.

Nº de inserciones	T O M		Metheringham		MARKSEL		Nº de casos
	Media	Desv. Tip.	Media	Desv. Tip.	Media	Desv. Tip.	
4	33,20%	1,77%	3,67%	1,29%	3,07%	1,07%	20
6	26,92%	2,23%	6,60%	2,27%	6,03%	2,20%	20
8	22,04%	2,82%	7,98%	1,54%	6,57%	1,63%	40
9	20,99%	1,65%	9,00%	2,60%	7,25%	1,76%	20
10	18,19%	1,47%	9,45%	2,32%	8,13%	1,89%	20
12	18,23%	2,10%	10,67%	2,10%	8,03%	1,95%	80
14	15,48%	1,87%	11,42%	2,19%	9,14%	1,94%	20
15	16,90%	2,24%	12,20%	1,96%	8,80%	1,95%	40
16	17,22%	2,09%	12,59%	1,98%	9,16%	2,02%	40
18	17,76%	2,78%	13,52%	1,56%	10,18%	1,81%	40
20	18,90%	3,28%	14,31%	2,20%	10,36%	2,01%	60
TODAS	19,61%	4,62%	10,92%	3,40%	8,31%	2,57%	400

Tabla 1.15: Errores en las distribuciones de contactos estimadas por TOM, Metheringham y MARKSEL (según el nº de inserciones).

En todos los niveles estima mejor el MARKSEL, luego Metheringham y después el TOM. Para cuatro inserciones, el TOM yerra la tercera parte de las distribuciones de frecuencias, diez veces más que los otros modelos. No conviene, por tanto, utilizar el TOM para estimar la distribución de contactos de pocas inserciones. Estas enormes diferencias se suavizan en las grandes campañas: Metheringham produce aproximadamente 3/4 partes de los errores del TOM, y el MARKSEL 1/2 de los mismos. La imprecisión de Metheringham crece más rápidamente que en MARKSEL. Otras investigaciones deberían analizar casos con más inserciones, donde quizá TOM mejore a Metheringham, pero muy probablemente nunca sea tan bueno como MARKSEL. Existe homogeneidad interna entre los grupos, pues las desviaciones típicas son pequeñas en relación a la media; además la variación de esas desviaciones no sigue ninguna tendencia con el nº de inserciones.

Nº de periodos	T O M		Metheringham		MARKSEL		Nº de casos
	Media	Desv. Tip.	Media	Desv. Tip.	Media	Desv. Tip.	
2	22,89%	5,96%	9,74%	4,10%	7,19%	2,99%	100
3	20,57%	4,00%	10,50%	3,26%	8,18%	2,30%	100
4	18,55%	2,97%	11,42%	2,95%	8,64%	2,32%	80
5	17,14%	2,02%	11,95%	3,10%	8,91%	2,36%	60
6	16,89%	2,07%	12,11%	2,54%	9,50%	2,11%	40
7	15,48%	1,87%	11,42%	2,19%	9,14%	1,94%	20
TODOS	19,61%	4,62%	10,92%	3,40%	8,31%	2,57%	400

Tabla 1.16: Errores en las distribuciones de contactos estimadas por TOM, Metheringham y MARKSEL (según el nº de periodos).

Cuando la campaña dura más periodos, TOM va mejorando sus estimaciones, mientras que los otros métodos van empeorando ligeramente (cfr. **Tabla 1.16**); era lo esperado pues el nº de inserciones depende directamente del nº de periodos. En dos periodos el TOM tiene sus mayores errores, y además la superior desviación típica indica su heterogeneidad; con el alcance sucedía exactamente lo contrario (**Tabla 1.13**): cuanto más simple sea la campaña, se estima mejor el alcance y peor la distribución de contactos.

Nº de soportes	T O M		Metheringham		MARKSEL		Nº de casos
	Media	Desv. Tip.	Media	Desv. Tip.	Media	Desv. Tip.	
2	22,03%	6,57%	8,27%	3,33%	6,91%	2,71%	120
3	17,93%	2,70%	11,35%	2,59%	8,53%	2,29%	80
4	18,92%	3,05%	11,57%	3,11%	8,72%	2,37%	80
5	17,64%	2,53%	13,34%	2,02%	9,80%	1,77%	40
6	19,14%	2,58%	11,81%	2,39%	8,49%	2,15%	40
8	17,08%	2,17%	12,49%	2,22%	8,73%	2,26%	20
10	21,98%	2,41%	14,23%	2,18%	10,48%	1,81%	20
TODOS	19,61%	4,62%	10,92%	3,40%	8,31%	2,57%	400

Tabla 1.17: Errores en las distribuciones de contactos estimadas por TOM, Metheringham y MARKSEL (según el nº de soportes).

El aumento de los soportes implica un zigzaguo en los errores del TOM, y un progresivo crecimiento en los demás (véase la **Tabla 1.17**). La influencia es mayor que

antes para Metheringham y MARKSEL, y menor para el TOM. En todos los niveles, MARKSEL sigue estimando mejor que Metheringham, y éste mejor que el TOM.

Hasta ahora hemos revisado la distribución de contactos en su conjunto, pero no **cada contacto**. Sin embargo, interesa también si un determinado contacto tiene desviaciones sistemáticas, si la forma de la curva se ajusta a la realidad, etc. La nueva medida del error es la diferencia entre la frecuencia estimada y observada del contacto correspondiente: $f'_i - f_i$. Se atiende a la diferencia simple y absoluta, para evaluar tanto el signo como el tamaño. La unidad de medida sigue siendo el porcentaje sobre el público objetivo. Los resultados se presentan resumidos en las **Tablas 1.18 y 1.19**, la primera para valores simples y la segunda para valores absolutos. En las filas se sitúan los errores medios, por grupos de campañas 4, 6, 8,..., 20 inserciones; en cada alternativas se comparan los tres modelos; por columnas están los errores medios por contacto que, como es obvio, sólo existen en los contactos menores al nº de inserciones.

El TOM ha subestimado el primer y el último contacto en todos los casos (error medio simple = error medio absoluto). La tendencia general es sobrestimar los demás contactos; al aumentar el nº de inserciones, la subestimación final tiende a empezar algún contacto antes. La subestimación media del primer contacto es mayor que la del último. La sobrestimación del segundo contacto es casi general, atenuándose cuando se incrementan las inserciones; además, con pocas inserciones, suele haber un máximo de las sobrestimaciones en el segundo contacto, pero se va desplazando a la derecha al aumentar las inserciones. Los errores del TOM siguen una curva en forma de sierra, con dos máximos absolutos, en el extremo izquierdo y en el extremo derecho.

El modelo de Metheringham produce errores de menor dimensión, porque la curva estimada busca el ajuste con la nube de frecuencias reales. En especial es relevante la **Tabla 1.18**, para analizar el signo: existe subestimación media aproximadamente en la primera cuarta parte de la distribución; también en la cuarta parte final, exceptuando que

en el último contacto se afianza la sobrestimación cuando aumentan las inserciones; en la mitad central existe la contrapuesta sobrestimación media. Estos resultados se deben a que Metherringham ha estimado distribuciones en forma de campanas (cóncavas hacia el eje de abcisas) ligeramente inclinadas hacia la derecha.

En el MARKSEL predominan ligeras subestimaciones medias en toda la distribución, que se compensan con la sobrestimación del último contacto. Los errores medios absolutos son menores que los del TOM, casi para cualquier contacto. La estabilidad de los errores, a lo largo de la distribución, es mayor que en Metherringham y, sobre todo, mayor que en el TOM.

Inser- ciones	MO- DELO	CONTACTOS																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	TOM Mether. MARK.	-9,35 -0,25 -0,59	9,91 1,18 -0,40	6,42 0,20 -0,35	-7,52 -0,77 0,52																
6	TOM Mether. MARK.	-8,07 -0,74 -0,20	5,15 0,79 0,00	3,01 0,57 -0,77	2,52 0,45 -0,38	2,26 -0,81 -0,43	-5,72 -0,32 0,75														
8	TOM Mether. MARK.	-6,86 -1,16 -0,16	2,88 -0,23 -0,25	2,11 0,46 -0,41	1,84 1,08 -0,08	1,24 0,70 -0,20	1,17 -0,37 -0,49	1,23 -1,08 -0,22	-4,28 0,30 0,99												
9	TOM Mether. MARK.	-6,53 -1,36 -0,26	2,58 -0,47 -0,17	1,49 0,14 -0,44	1,30 0,65 -0,36	1,78 1,49 0,47	0,76 0,37 -0,26	1,27 -0,38 -0,21	0,52 -1,53 -0,58	-3,84 0,52 0,98											
10	TOM Mether. MARK.	-5,83 -0,98 0,11	2,15 -0,76 -0,32	1,52 0,23 -0,18	0,67 0,44 -0,50	0,91 0,80 -0,30	0,66 0,74 -0,20	0,81 0,35 -0,11	0,69 -1,01 -0,74	0,14 -1,26 -0,36	-2,72 0,80 1,51										
12	TOM Mether. MARK.	-5,33 -1,22 -0,17	1,66 -0,86 -0,08	1,22 -0,31 -0,26	0,91 0,66 0,14	0,54 0,54 -0,28	0,64 0,67 -0,24	0,58 0,80 -0,02	0,57 0,39 -0,15	0,54 -0,05 -0,09	0,62 -0,95 -0,33	0,41 -0,97 0,01	-2,99 0,54 0,74								
14	TOM Mether. MARK.	-4,39 -0,79 0,09	0,81 -1,05 -0,29	1,14 -0,46 -0,20	0,73 0,23 -0,01	0,01 -0,03 -0,62	0,48 0,86 0,06	0,55 0,82 -0,05	0,23 0,70 -0,10	0,38 0,31 -0,29	0,42 0,17 -0,10	0,02 -0,57 -0,40	0,74 -0,93 -0,30	-0,27 -0,79 0,04	-1,83 0,71 1,13						
15	TOM Mether. MARK.	-4,26 -0,71 0,12	1,07 -1,10 -0,15	0,95 -0,77 -0,30	0,42 -0,05 -0,09	0,40 0,35 -0,06	0,36 0,46 -0,15	0,64 0,81 0,10	0,49 0,68 -0,04	0,20 0,57 -0,06	0,63 0,44 0,00	-0,03 -0,14 -0,25	0,33 -0,62 -0,27	0,64 -1,05 -0,23	-0,18 -0,76 0,07	-2,20 0,85 0,67					
16	TOM Mether. MARK.	-3,92 -0,48 0,28	0,85 -1,29 -0,27	1,23 -0,58 0,01	0,54 -0,08 -0,01	0,08 0,12 -0,19	0,30 0,52 -0,02	0,46 0,61 -0,04	0,42 0,47 -0,21	0,23 0,59 -0,05	0,39 0,46 -0,07	0,59 0,56 0,24	0,06 -0,10 -0,09	0,59 -0,56 -0,09	0,49 -1,28 -0,39	-0,49 -0,76 0,01	-2,08 0,92 0,55				
18	TOM Mether. MARK.	-3,59 -0,45 0,14	0,65 -1,07 -0,10	0,82 -0,90 -0,19	0,65 -0,36 -0,09	0,26 0,15 0,03	0,36 0,64 0,26	0,20 0,34 -0,20	0,08 0,32 -0,30	0,36 0,58 -0,06	0,16 0,47 -0,14	0,17 0,42 -0,12	0,56 0,63 0,24	-0,02 -0,02 -0,19	0,18 -0,38 -0,22	0,06 -0,78 -0,20	1,10 -0,93 -0,04	-0,80 -0,56 0,10	-1,62 0,90 0,59		
20	TOM Mether. MARK.	-3,24 -0,28 0,13	0,34 -1,14 -0,19	0,72 -0,89 -0,06	0,98 -0,42 0,02	0,01 -0,05 0,01	0,21 0,06 -0,15	-0,07 0,43 0,03	0,46 0,56 0,07	0,32 0,53 -0,01	0,40 0,46 -0,10	-0,04 0,46 -0,09	0,27 0,49 0,01	0,17 0,29 -0,10	0,51 0,35 0,13	-0,14 0,04 0,05	0,53 -0,56 -0,24	-0,42 -0,88 -0,19	1,34 -1,11 -0,22	-1,30 -0,39 0,11	-1,36 0,97 0,43

en %

Tabla 1.18: Errores en los contactos estimados por TOM, Metheringham y MARKSEL (medias según el n° de inserciones).

Inser- ciones	MO- DELO	CONTACTOS																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	TOM Mether. MARK.	9,35 0,74 0,81	9,91 1,28 0,90	6,42 0,65 0,64	7,52 1,01 0,72																
6	TOM Mether. MARK.	8,07 1,03 0,86	5,15 1,00 0,75	3,15 1,14 1,13	2,52 1,28 1,08	2,29 1,40 1,32	5,72 0,75 0,89														
8	TOM Mether. MARK.	6,86 1,19 0,65	2,88 0,81 0,78	2,11 0,90 0,83	1,84 1,11 0,67	1,49 1,05 0,86	1,23 0,86 0,92	1,34 1,19 0,74	4,28 0,89 1,13												
9	TOM Mether. MARK.	6,53 1,40 0,72	2,58 0,83 0,75	1,59 0,74 0,82	1,30 0,75 0,74	1,78 1,53 0,87	1,07 0,82 0,64	1,36 0,73 0,75	0,93 1,53 0,90	3,84 0,67 1,06											
10	TOM Mether. MARK.	5,83 1,02 0,55	2,15 0,97 0,74	1,52 0,47 0,52	1,22 0,94 0,88	1,11 1,06 0,80	0,98 1,02 0,74	0,95 0,70 0,70	0,99 1,06 0,91	0,72 1,30 0,77	2,72 0,92 1,51										
12	TOM Mether. MARK.	5,33 1,25 0,60	1,68 1,02 0,67	1,32 0,78 0,76	1,03 0,80 0,55	0,92 0,88 0,80	0,92 0,85 0,70	0,82 0,96 0,61	0,93 0,75 0,68	0,75 0,63 0,63	0,84 1,04 0,65	0,69 1,04 0,57	2,99 0,69 0,82								
14	TOM Mether. MARK.	4,39 0,79 0,44	0,87 1,05 0,55	1,14 0,70 0,64	1,05 0,79 0,76	0,66 0,64 0,86	0,78 0,88 0,45	0,63 0,88 0,43	0,63 0,77 0,62	0,80 0,74 0,76	0,71 0,83 0,73	0,59 0,74 0,63	0,85 0,97 0,61	0,56 0,81 0,52	1,83 0,83 1,13						
15	TOM Mether. MARK.	4,26 0,76 0,43	1,14 1,18 0,60	1,11 0,89 0,64	0,71 0,55 0,54	0,79 0,59 0,50	0,69 0,69 0,61	0,89 0,95 0,65	0,79 0,86 0,61	0,81 0,83 0,67	0,78 0,63 0,52	0,69 0,52 0,56	0,71 0,79 0,63	0,82 1,12 0,64	0,51 0,85 0,47	2,20 0,98 0,73					
16	TOM Mether. MARK.	3,92 0,57 0,47	0,97 1,32 0,72	1,25 0,77 0,56	0,78 0,57 0,58	0,77 0,67 0,68	0,82 0,76 0,71	0,76 0,83 0,59	0,88 0,73 0,66	0,68 0,77 0,49	0,75 0,72 0,57	0,72 0,63 0,49	0,66 0,57 0,54	0,67 0,65 0,46	0,83 1,29 0,67	0,66 0,83 0,38	2,08 0,92 0,57				
18	TOM Mether. MARK.	3,59 0,60 0,46	0,82 1,12 0,60	1,03 1,04 0,73	0,90 0,66 0,60	0,65 0,52 0,48	0,71 0,75 0,56	0,60 0,65 0,56	0,75 0,69 0,65	0,76 0,71 0,53	0,69 0,64 0,55	0,75 0,73 0,64	0,84 0,76 0,58	0,75 0,63 0,65	0,64 0,68 0,61	0,65 0,87 0,57	1,15 0,95 0,41	0,85 0,63 0,39	1,62 0,91 0,60		
20	TOM Mether. MARK.	3,24 0,50 0,43	0,67 1,15 0,50	0,88 0,94 0,55	1,12 0,66 0,54	0,65 0,48 0,48	0,66 0,60 0,59	0,66 0,63 0,48	0,73 0,74 0,56	0,63 0,71 0,51	0,72 0,66 0,52	0,68 0,67 0,59	0,72 0,74 0,63	0,62 0,58 0,50	0,67 0,60 0,53	0,64 0,45 0,46	0,74 0,70 0,54	0,80 0,90 0,52	1,38 1,13 0,51	1,32 0,51 0,40	1,36 0,97 0,52

en %

Tabla 1.19: Errores absolutos en los contactos estimados por TOM, Metherringham y MARKSEL (medias según el nº de inserciones).

PARTE II

LA AUDIENCIA NETA

INTRODUCCIÓN

Bajo el título de la **audiencia neta**, en esta Parte se estudian el alcance y la distribución de contactos de múltiples soportes en un momento determinado. El alcance neto es el número de individuos del público objetivo que están expuestos al menos a uno de esos soportes; la cobertura es el porcentaje de individuos del público objetivo alcanzado por esos soportes. Hemos querido mantener los términos *audiencia neta*, *audiencia acumulada* y *audiencia neta acumulada*, tal como se han acuñado durante años en la literatura vernácula⁴⁴. Se van a analizar en tres Partes, que se diferencian en acometer el único objeto de esta tesis desde diversos puntos de vista. Nótese que no distinguimos la *audiencia útil* (parte de la audiencia total que está incluida en el público objetivo) porque las audiencias siempre se miden sobre los "targets" específicos.

Se exponen primero los modelos ad-hoc y luego los estocásticos, clasificación realizada por Chandon⁴⁵; los modelos de simulación serán tratados después. La presentación ordinariamente está estructurada en tres partes: exposición del modelo en sí, ejemplificación y aplicación a un estudio empírico en nuestro país. La exposición teórica va acompañada a veces de investigaciones posteriores a cargo de otros autores, así como de la adaptación del planteamiento a España. El ejemplo, cuyo objetivo es facilitar la comprensión de las formulaciones teóricas, es uno caso sencillo del estudio empírico. Esta investigación empírica se basa en comparar el alcance neto estimado por los modelos con el alcance neto obtenido por el programa TOM2, para 402 campañas; lo mismo se hace respecto a las distribuciones de contactos. Las audiencias del EGM están probabilizadas, pues la encuesta mide el hábito del entrevistado, en vez del

⁴⁴ Véase:

- MARTÍN ARMARIO, Enrique. "La gestión publicitaria". Ed. Pirámide. Madrid, 1980. Págs. 151-8.
- MARTÍNEZ RAMOS, Emilio. "El uso de los medios de comunicación en marketing y publicidad". Ed. Akal/Comunicación. Madrid, 1992. Págs. 54-93.
- DÍEZ DE CASTRO, Enrique Carlos y MARTÍN ARMARIO, Enrique. "Planificación publicitaria". Ed. Pirámide. Madrid, 1993. Págs. 277-335.

⁴⁵ CHANDON, Jean-Louis. "Le point sur les modèles de duplication d'audience". Revue Française du Marketing. Nº 53. 1974. Págs. 51-76.

comportamiento puntual; la forma de los datos cambia respecto a otros autores y se hace más necesaria la verificación empírica a partir de las fuentes disponibles en España; de esta forma, los modelos podrán ser sustitutivos del TOM, utilizando datos del EGM. Los fundamentos de esa investigación aparecen en el primer modelo, la fórmula de Agostini (§ B.1.4).

Finalmente se hace un análisis comparativo entre modelos, ad-hoc y estocásticos, llegándose a algunas conclusiones acerca de la precisión de su estimación para el caso español.

A) MODELOS AD-HOC

Conceptos preliminares

Los **modelos ad-hoc, deterministas o empíricos** son aquellos que determinan la audiencia neta a partir de fórmulas obtenidas por análisis de regresión u otros procedimientos adecuados. Se caracterizan por la ausencia de hipótesis y por su naturaleza no estocástica o aleatoria. Además, a diferencia de los modelos estocásticos, no permiten estimar la distribución de contactos.

Se recogen aquí siete modelos ad-hoc, ordenados según su momento de aparición. A la pionera fórmula de *Agostini* le siguen dos alternativas, presentadas por *Kuhn* y *Kaatz*, para mejorar sus estimaciones. Después aparece una reformulación a cargo de *Hofmans*. En quinto lugar tenemos el modelo de *Starch*, con distintos fundamentos. Más tarde *Little* y *Lodish* hicieron un completo modelo de simulación, pero en esta Parte se estudiará exclusivamente la duplicación entre parejas de soportes. Y finalmente el español *Lozano* propuso un sencillo modelo a partir de los datos del EGM.

1. La fórmula de Agostini

1.1. PLANTEAMIENTO DEL AUTOR

Los estudios elaborados por **Agostini**⁴⁶ concluyeron en un método empírico de cálculo, que permite explicar el alcance neto de una combinación de soportes, en medios impresos, con una sola inserción por soporte. La investigación se basa en las encuestas realizadas cuatro años antes por Sema para el Centre d'Etude des Supports de Publicité (CESP).

Esas encuestas se referían a 15 revistas francesas. El número de diferentes combinaciones que se podían hacer con 15 revistas era enorme: $2^{15} = 32.768$. Se publicaba el alcance neto de todas, es decir, 105 combinaciones de 2 revistas, 455 de 3 revistas, 1.365 de 4 revistas, 3.003 de 5 revistas,..., y 1 de 15 revistas. Con un número relativamente pequeño de revistas, el número de combinaciones es bastante elevado; pero si se consideraran 100 soportes, habría 2^{100} posibles combinaciones.

El alcance neto de una combinación de n soportes se podría calcular a partir de los que leen simultáneamente 2, 3, 4,..., ó n soportes (duplicaciones, triplicaciones,...). Sin embargo, la fórmula descubierta por **Agostini** permite hallar el alcance neto de cualquier combinación de soportes, a partir únicamente de las audiencias y de las duplicaciones entre parejas de soportes.

Las encuestas del CESP recogían también las duplicaciones de 15 revistas adicionales, pero no todas las posibles combinaciones de ellas. Se contaba, por tanto, con las duplicaciones de 30 revistas. La fórmula de Agostini permitía conocer el alcance neto de una campaña que incluyera esas revistas cuyos únicos datos disponibles eran las duplicaciones.

⁴⁶AGOSTINI, Jean-Michel. "How Estimate Unduplicated Audiences". *Journal of Advertising Research*. Marzo 1961. Vol 1. Nº 3. Págs. 11-14.

Teniendo en cuenta la hipótesis de que se realiza una sola inserción por soporte, supongamos una campaña publicitaria que se va a insertar en n soportes identificados como $S_1, S_2, S_3, \dots, S_n$, cuyas audiencias individuales son $A_1, A_2, A_3, \dots, A_n$.

Las duplicaciones de audiencia para los n soportes, tomados de dos en dos, se muestran en la **Tabla 2.1**. La intersección de las respectivas filas y columnas de los soportes, indica la población que simultáneamente está expuesta al soporte de la fila y al soporte de la columna.

<i>SOPORTES</i>	S_1	S_2	S_3	S_4	...	S_{n-1}	S_n
S_1		D_{12}	D_{13}	D_{14}	...	$D_{1,n-1}$	D_{1n}
S_2			D_{23}	D_{24}	...	$D_{2,n-1}$	D_{2n}
S_3				D_{34}	...	$D_{3,n-1}$	D_{3n}
.
.
.
S_{n-1}					...		$D_{n-1,n}$
S_n					...		

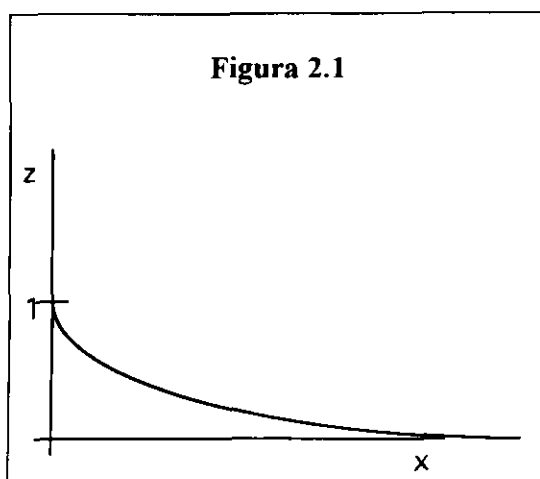
Tabla 2.1

Se llama D_{ij} a las duplicaciones de audiencia de los soportes S_i y S_j ; por Combinatoria sabemos que existen $n(n-1)/2$ posibles duplicaciones. Se ha rellenado sólo la parte superior derecha de la tabla, ya que las duplicaciones de un soporte con otro son iguales a las de este último con aquél; se cumple por tanto la igualdad: $D_{ij} = D_{ji}$. Se trata de una matriz simétrica. No se han rellenado las combinaciones de los soportes consigo mismos (diagonal principal de la matriz) ya que partíamos de la hipótesis de única inserción por soporte.

El alcance neto (C) será menor que la suma de las audiencias de los vehículos (A) o, en caso extremo, será igual si los soportes tienen audiencias excluyentes. Luego la relación $C \leq A$ puede ser expresada en forma de igualdad introduciendo un factor Z tal que

$$C = Z \cdot A \quad (0 < Z \leq 1)$$

Introdujo una nueva variable x , resultado de dividir la suma de las duplicaciones por la suma de audiencias simples (audiencia bruta): $x = \frac{D}{A}$. Descubrió que entre Z y x existía una relación funcional sencilla que, gráficamente, puede ser expresada por una curva, de forma tal que para cada valor de x , existirá un valor de Z , que llevado a $C = Z \cdot A$, nos permitirá calcular el alcance de los n soportes. Dicha función se muestra en la **Figura 2.1**.



Si suponemos que no hay duplicación entre los soportes ($D=0$), Z tendría que tomar el valor 1, ya que $C=A$. Nos encontraríamos en el punto $(0,1)$ de la gráfica. Por otra parte, cuando x crece, Z decrece, ya que a mayor duplicación menor alcance neto; luego la curva seguirá la forma de la figura, conservándose siempre en el primer cuadrante, único en el que la función tiene significado. Agostini dibujó 98 puntos, con sus respectivos valores de x y Z . Cada punto correspondía a una combinación de soportes, elegida aleatoriamente entre las 32.767 posibles combinaciones de las 15 revistas recogidas por el CESP. Procedió después al ajuste de esa nube de puntos, obteniendo que la curva que mejor ajustaba era una hipérbola equilátera, asintótica con el eje x , de ecuación $axZ + bZ + c = 0$ ⁴⁷.

En el punto $(0,1)$ la ecuación cumple que $b = -c$. Luego

⁴⁷ SANZ DE LA TAJADA, Luis Ángel y RUBIRA BLÁZQUEZ, Francisco. "Tecnología de los medios publicitarios".

$$axZ + bZ - b = 0 \quad \Rightarrow \quad \frac{a}{b}xZ + Z - 1 = 0.$$

Llamando $\frac{a}{b} = k$,

$$kZx + Z - 1 = 0 \quad \Rightarrow \quad Z = \frac{1}{k \cdot x + 1}$$

Sustituyendo $Z = \frac{C}{A}$ y $x = \frac{D}{A}$, y despejando C, se llega a la fórmula conocida como de Agostini:

$$C = \frac{A}{k \frac{D}{A} + 1}$$

siendo,

C, el alcance neto;

$A = \sum_{i=1}^n A_i$, la suma de audiencias de los n soportes; $i=\{1,2,\dots,n\}$.

$D = \frac{1}{2} \left[\sum_{i=1}^n D_{ij} - \sum_{i=1}^n D_{ii} \right]$, la suma de todas las duplicaciones⁴⁸ distintas de los n

soportes, tomados dos a dos, con las condiciones antes impuestas; $i=\{1,2,\dots,n\}$; $j=\{1,2,\dots,n\}$.

k , constante que toma el valor de 1'125.

Al tratarse de una fórmula obtenida empíricamente, sin derivarse de premisas más generales, se debe plantear con datos distintos que verifiquen la fórmula. Agostini señalaba que el parámetro k tomaba el valor 1'125 cuando se cumplían las hipótesis

⁴⁸ SANZ DE LA TAJADA, Luis Ángel. "La publicidad: una técnica al servicio de la acción comercial". Ediciones ESIC. Madrid, 1975. Pág. 204.

empleadas en su investigación; no estaba planteado una fórmula general para todos los medios y todos los países. Se estaba abriendo una línea de investigación (como señaló Ramond⁴⁹ en el mismo número de la revista) que produciría abundante literatura.

1.2. EJEMPLO

Hemos escogido como ejemplo una de las campañas planificadas en esta tesis, según se explica posteriormente (§A.1.4). Ahora sólo queremos facilitar la comprensión de la fórmula anterior, como haremos con todos los modelos.

Se dirige a un público objetivo formado por hombres, de 14 a 34 años, de clase social alta y media alta. En España existen 1.455.001 personas que cumplen esas condiciones, según los datos del EGM. Se han seleccionado los cuatro soportes con mayor audiencia en ese *target*, haciendo una inserción por soporte. Corresponde a la campaña nº 6 del fichero **ALCANCE.WK1**, en el diskette anejo. Aquí cogemos la que tiene menos revistas para simplificar el ejemplo. Las revistas son Muy Interesante, Teleprograma, El Jueves y Teleindiscreta. En la **Tabla 2.2** figuran las audiencias de esas revistas (en negrilla), y la duplicación de audiencia entre el soporte fila y el soporte columna.

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.2

Calculamos la suma de audiencias (diagonal principal de la tabla) y la suma de las seis posibles duplicaciones:

⁴⁹ RAMOND, Charles K. "Agostini's Constant". *Journal of Advertising Research*. Marzo 1961. Vol 1. Nº 3.

$$A = \sum_{i=1}^4 A_i = 309.918 + 236.902 + 130.265 + 129.992 = 807.077$$

$$D = 55.208 + 47.445 + 30.036 + 23.935 + 29.574 + 11.555 = 197.753$$

Ya estamos en condiciones de hallar el alcance neto, partiendo de la hipótesis de que k vale 1'125 como en otros países:

$$C = \frac{A}{k \frac{D}{A} + 1} = \frac{807.077}{1,125 \cdot \frac{197.753}{807.077} + 1} = 632.678$$

1.3. INVESTIGACIONES POSTERIORES

1.3.1. Aplicaciones en Estados Unidos y Canadá

Bower⁵⁰ investigó el modelo de Agostini sobre revistas de Estados Unidos y Canadá. Se trataba de testar dicha fórmula a partir de siete estudios, con un total de 640 campañas combinando revistas representativas de los dos países. Estrictamente los estudios no son comparables y debían considerarse separadamente.

El primer análisis evaluaba el error de estimación (diferencia entre los valores del alcance neto real y estimado por la fórmula de Agostini), expresado en porcentaje. La media de este porcentaje, para los siete estudios, era de un 3,1%. De esta forma, se puede aceptar la fórmula para las revistas de Estados Unidos y Canadá.

Pretendía conocer también si la fórmula de Agostini se comportaba igual con distinto número de vehículos o con revistas no homogéneas. Con datos tomados del Behavior Systems, simuló 35 campañas con 3, 4, 5, 6 y 7 inserciones; unas campañas sólo para revistas médicas de información general; otras para revistas médicas especializadas; y otras para diversas combinaciones entre información general y especializadas.

Otro intento era calcular el coeficiente de correlación entre el error y la desviación estándar relativa de las duplicaciones dos a dos. Usando una muestra aleatoria de 100 casos, se obtuvo que el coeficiente de correlación r era 0,61 (que manifiesta una relación positiva con cercana certeza), y el coeficiente de determinación r^2 era de 0,37 (que indica que algo más de la tercera parte de la variación del error estimado puede atribuirse a la variación de la desviación estándar relativa de las duplicaciones dos a dos).

⁵⁰ BOWER, John. "Net Audiences of U.S. and Canadian Magazines: Seven Test of Agostini's Formula". *Journal of Advertising Research*. Marzo, 1963. Vol 3. Nº 1. Págs. 13-20.

Las conclusiones obtenidas por **Bower** se pueden resumir en cuatro puntos:

1. La técnica de Agostini parece ser generalmente aplicable, conclusión obtenida tras conocer los porcentajes de error de 640 casos.

2. Parece que al incrementarse el número de inserciones de la campaña, se incrementa el error. Pero no se ha podido probar ya que hay muy pocos casos que combinen una buena cantidad de revistas. Si finalmente se demostrara, se podría introducir un factor de corrección cuando se utilizan muchos soportes.

3. Los soportes no pueden determinarse aleatoriamente, como hizo Agostini, ya que se ha detectado una correlación positiva entre el error estimado y las parejas de duplicaciones. Los vehículos deberían ser homogéneos.

4. La fórmula de Agostini produce sobrestimaciones de la audiencia neta en el 90% de los casos, de manera que la constante $k = 1,125$ parece demasiado pequeña.

1.3.2. Aplicaciones en el Reino Unido

Caffyn y Sagovsky⁵¹ establecieron en el Reino Unido una comparación entre la fórmula de Agostini y los métodos de Sainsbury (§B.2 y B.3) en un nuevo medio: la Prensa Diaria. Comentan los autores que este medio tiene características similares al de revistas de información general en Francia o Estados Unidos. Los datos están tomados del IPA National Readership Surveys, que publica periódicamente en el Reino Unido las audiencias y duplicaciones.

Se trataba de estudiar las 42 posibles combinaciones de 6 periódicos: 20 de tres vehículos, 15 de cuatro vehículos, 6 de cinco vehículos y 1 con los seis. Los periódicos son diarios, excepto uno que es dominical. Los modelos que mejor han estimado la

⁵¹ CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods". *Journal of Advertising Research*. Marzo 1963. Vol. 3. Nº 1. Págs. 21-25.

audiencia neta de esas combinaciones son: Agostini, en 14 ocasiones; Sainsbury Modificado, en 21; Sainsbury Simple, en 10. Suman más de 42 porque en caso de empate se ha contado en ambos métodos. Si olvidáramos el Sainsbury Simple, Agostini ajustaría mejor en 15 casos y Sainsbury Modificado en 28.

La fórmula de Agostini estimaba mejor el alcance neto cuando todos los soportes eran periódicos diarios pero, cuando se consideraba el dominical, siempre era mejor el método de Sainsbury Modificado.

Caffyn y Sagovsky concluyen que existe una relación entre el valor x ($x = D/A$) de la fórmula de Agostini y sus errores de estimación. Para minimizar el error habría que incrementar el factor $kx+1$ cuando se incrementa el valor de x : dentro de la estructura de la fórmula esto implica un cambio en el valor de k . También apuntan que el valor de k debería cambiar ligeramente según el número de vehículos utilizados.

1.3.3. Aplicaciones en Francia

Marc⁵² intentó aplicar la fórmula de Agostini a revistas especializadas de negocios en Francia. Este es un medio donde la mayoría de los lectores leen tres o cuatro revistas al mismo tiempo, no por diversión sino por necesidad. La duplicación será, por tanto, particularmente alta y habrá que comprobar la validez de la fórmula.

Se realizó una encuesta entre los fabricantes de muebles sobre la audiencia de cinco revistas de muebles. Se determinaron todas las posibles duplicaciones, triplicaciones, cuadruplicaciones y la quintuplicación: en total 26 combinaciones. Para cada uno de esos casos, se determinó x y Z , obteniéndose una hipérbola equilátera similar a la de Agostini, pero con distinto coeficiente: k tomaba el valor de 1,6.

⁵² MARC, Marcel. "Net Audiences of French Business Papers: Agostini's Formula Applied to Special Markets". *Journal of Advertising Research*, Marzo 1963, Vol. 3, Nº 1. Págs. 26-29.

Manteniendo la tesis -no demostrada- de que la constante k varía según la estructura de la audiencia de las revistas seleccionadas, Marc había planteado⁵³ un mecanismo para determinar dicho parámetro:

$$k = \frac{(L_2 + 2L_3 + \dots + (n-1)L_n) \cdot (L_1 + 2L_2 + \dots + nL_n)}{(L_1 + L_2 + \dots + L_n) \cdot (L_2 + 3L_3 + 6L_4 + \dots + C_2^n L_n)}$$

siendo,

L_1 : lectores de una sola revista.

L_2 : lectores de dos revistas.

L_3 : lectores de tres revistas.

L_n : lectores de las n revistas consideradas.

C_2^n : número de posibles combinaciones de n revistas tomadas de dos en dos (sin repetición).

Cada lector pertenece únicamente a un L_i . El sumatorio de todos los L_i constituye el total de individuos de la muestra, restados los que no leen ninguna revista.

Se realizaron cuatro encuestas para testar estas hipótesis:

- a) la ya comentada de revistas de muebles;
- b) una de revistas económicas leídas por hombres de negocios; se determinó que k tomaba el valor de 1,436.
- c) revistas de información general leídas por hombres de negocios; $k = 1,318$.
- d) revistas de limpieza leídas por trabajadores de estos servicios; $k = 1,510$.

Marc concluye que con su método se puede determinar el alcance neto, conociendo la audiencia de todos soportes, las duplicaciones dos a dos y la distribución de la población de acuerdo con el número de revistas leídas. También dice que no se trata más que de una generalización de la fórmula de Agostini a revistas económicas, o a revistas de información general leídas por hombres de negocios. Se trataría ahora de

⁵³ MARC, Marcel. "Étude de la Couverture Réalisée par Plusieurs Revues Techniques sur une Clientèle Spécialisée". Institut de Recherches et d'Études Publicitaires, estudio nº6. Noviembre 1962.

determinar el valor de k para cada uno de los grupos de audiencia: de un país o de otro, con una edad determinada, ricos o pobres, en la ciudad o en el campo,...

1.3.4. Aportación de Claycamp y McClelland

Claycamp y McClelland⁵⁴ hicieron una revisión del modelo de Agostini, con tres objetivos:

- a) dar una explicación analítica del valor de k y de las variaciones en la exactitud de sus estimaciones.
- b) mostrar cómo se puede estimar un valor apropiado de k para una base de datos determinada.
- c) discutir la constancia del parámetro k .

La interpretación analítica de k parte de la noción de frecuencia o número medio de soportes con que contacta un lector (se mantiene todavía la hipótesis una sola inserción por soporte). La frecuencia será, por tanto, el cociente entre el alcance bruto y el alcance neto, y la notaremos como μ ; con la terminología anterior, se llega a una nueva formulación del modelo de Agostini:

$$\frac{A}{C} = \mu \quad \Leftrightarrow \quad \mu = 1 + k \cdot \frac{D}{A}$$

En esa expresión se reconoce el *significado de k* : mide la relación entre D/A y la frecuencia. Claycamp y McClelland plantean después la variación de k a partir del análisis de la distribución de frecuencias de los lectores de n soportes. Define N_i como el número de personas que leen exactamente i soportes. La suma de todas esas personas será el alcance neto, mientras que esa suma multiplicada por los correspondientes soportes leídos será el alcance bruto:

⁵⁴ CLAYCAMP, H.J. and MCCLELLAND, C.W. "Estimating Reach and the Magic of K ". *Journal of Advertising Research*. Junio 1968. Vol. 8, Nº 2. Págs. 44-51.

$$C = \sum_{i=1}^n N_i \qquad A = \sum_{i=1}^n i \cdot N_i$$

Llamando $f(i)$ a la distribución de frecuencias que describe la probabilidad de que un individuo, seleccionado aleatoriamente de la muestra, lea exactamente i soportes, la frecuencia N_i será una fracción del alcance neto:

$$N_i = f(i) \cdot C$$

Se puede expresar ahora el alcance bruto en función del alcance neto, donde μ es la media de la distribución de frecuencias:

$$A = \sum_{i=1}^n i \cdot f(i) \cdot C = C \cdot \sum_{i=1}^n i \cdot f(i) = C \cdot \mu$$

La suma de duplicaciones (D) se puede definir en función de la varianza y de la media de la distribución de frecuencias:

$$\begin{aligned} D &= \sum_{i=1}^n \binom{i}{2} N_i = \sum_{i=1}^n \frac{i \cdot (i-1)}{2} \cdot f(i) \cdot C = \frac{C}{2} \cdot \sum_{i=1}^n (i^2 - i) \cdot f(i) = \\ &= \frac{C}{2} \left[\sum_{i=1}^n i^2 \cdot f(i) - \sum_{i=1}^n i \cdot f(i) \right] \end{aligned}$$

Como la varianza (σ^2) se define como la suma de cuadrados menos la media al cuadrado:

$$\sigma^2 = \sum_{i=1}^n i^2 \cdot f(i) - \mu^2$$

Sustituyendo en la ecuación anterior, llegaremos finalmente al valor de la suma de las duplicaciones:

$$D = \frac{C}{2} \cdot \left[\sigma^2 + \mu^2 - \sum_{i=1}^n i \cdot f(i) \right] = \frac{C}{2} \cdot [\sigma^2 + \mu^2 - \mu]$$

Remplazando D en la fórmula de Agostini, se consigue definir k en función de la media y varianza de la distribución de frecuencias:

$$\mu = 1 + k \cdot \frac{D}{A} = 1 + k \cdot \frac{\frac{C}{2} \cdot [\sigma^2 + \mu^2 - \mu]}{A} \Leftrightarrow$$

$$k = \frac{(\mu - 1) \cdot 2 \cdot A}{[\sigma^2 + \mu^2 - \mu] \cdot C} \Leftrightarrow k = \frac{2\mu(\mu - 1)}{\mu(\mu - 1) + \sigma^2}$$

De esa ecuación se obtienen tres conclusiones:

$$a) \lim_{\mu \rightarrow 0} k = 0$$

$$b) \lim_{\mu \rightarrow n} \sigma^2 = 0$$

$$c) \lim_{\mu^2 \rightarrow 0} k = 2$$

Luego se cumplirá que $0 \leq k \leq 2$. Además ese parámetro tomará su mayor valor ($k=2$) cuando todos los individuos lean exactamente r soportes ($1 \leq r \leq n$). Un caso especial se plantea cuando $n=2$; entonces $k=\mu$, como se demuestra al sustituir la duplicación ($D=A-C$) en la fórmula de Agostini:

$$\mu = 1 + k \cdot \frac{D}{A} = 1 + k \cdot \frac{A - C}{A} \Leftrightarrow$$

$$k = (\mu - 1) \cdot \frac{A}{A - C} = \frac{A - C}{C} \cdot \frac{A}{A - C} = \frac{A}{C} = \mu$$

Claycamp y McClelland investigaron la naturaleza de la variación de k y su efecto en la *estimación del alcance neto*. Para ello examinaron 28 revistas técnicas, a partir de una muestra de 491 ingenieros y científicos, en Estados Unidos. Primeramente quieren estimar k para 100 campañas, eligiendo su número de inserciones aleatoriamente en cada caso. Los resultados mostraban que la estimación de k era significativamente diferente, haciéndose mayor al incrementarse D/A y μ . Se producía una subestimación de C/A para pequeños valores de D/A y, por el contrario, una sobrestimación para grandes valores.

En segundo lugar estudian otras 100 campañas, divididas en cuatro conjuntos de 3, 5, 10 y 20 inserciones. Los resultados reflejaban que el valor de k aumentaba con el número de inserciones, y la varianza de k se reducía. Por otra parte, los resultados mostraban que la variación de k ha sido bastante insensible ante el tipo de soporte, especialmente si el número de inserciones era elevado. Para explicar la buena estimación de la fórmula de Agostini, aunque no se verifique la constancia de k , estudian matemáticamente la variación del alcance neto ante variaciones de k :

$$C/A = \frac{1}{1 + k \cdot D/A}$$

$$\frac{\partial(C/A)}{\partial k} = - \frac{D/A}{(1 + k \cdot D/A)^2} \cdot \partial k$$

Al cumplirse siempre que $\frac{D/A}{(1 + k \cdot D/A)^2} < 1$, se puede deducir que la variación de C/A será siempre menor que la variación de k . Empíricamente se muestra además que es considerablemente menor.

Como conclusión final, Claycamp y McClelland afirman que si k ha sido estimado específicamente para una base de datos, los errores que se derivan de tomarlo como un parámetro constante pueden ser ignorados.

1.4. INVESTIGACIÓN REALIZADA EN NUESTRO PAÍS

El primer objetivo era conocer el valor en España de la constante k . Siguiendo el método empírico utilizado por Agostini, se planificaron 402 campañas publicitarias (cuatro veces más que las del publicitario francés) insertadas en 30 revistas (dos veces más). Para ello, conté con la colaboración del **Estudio General de Medios** que me proporcionó los datos de audiencia recogidos periódicamente por sus encuestas y, después, con la empresa **ODEC** para el tratamiento informático de dicha información mediante los programas TOM. A ellos mi agradecimiento por su colaboración generosa.

El análisis se ha realizado únicamente sobre el Medio Revistas. Se pretendía que las conclusiones de este estudio se pudieran generalizar para todo el Medio. Para ello se debió seleccionar una muestra que fuera representativa del total. En España no se conoce el número exacto de revistas⁵⁵: antiguamente era obligatorio inscribirlas en un registro antes de su lanzamiento, pero no se inscribía su cese en la actividad; después se quitó esta obligatoriedad para las revistas que no solicitaran la ayuda del Estado; entonces, a falta de algún censo o fuente fiable, no se sabe cuántas revistas se publican. Se las suele dividir en revistas de información general y en publicaciones especiales; las primeras contrataban publicidad en 1990 por valor de algo más de 50.000 millones, mientras que se estima que las publicaciones especiales contratarían un poco menos de 70.000 millones; eso supone que las revistas de información general alcanzan apenas un 40% del total de la inversión publicitaria en el Medio Revistas⁵⁶.

Para seleccionar la muestra nos servimos del programa **TOM1** (Clasificación de Soportes), que permite ver, para un público objetivo dado, una selección de soportes según la cobertura, la rentabilidad o la afinidad, que obtengan (véase §C.1).

⁵⁵ GONZÁLEZ LOBO, M^a Ángeles. "Curso de Publicidad". Ed. Eresma & Celeste Ediciones. Madrid. 1994. Pág. 225.

⁵⁶ J. WALTER THOMPSON. "La inversión publicitaria en España: 1990". Madrid. 1991. Págs. 58-59.

Los datos utilizados fueron los correspondientes a las encuestas realizadas por el Estudio General de Medios durante el año 1991, considerando el acumulado de las tres olas para obtener mayor precisión (cfr. ficha técnica en **Figura 1.3, §B**). El público objetivo no estaba sujeto a ninguna variable.

La salida de ordenador del programa TOM1 nos mostraba la lista de las 86 revistas controladas por ODEC. Por audiencia de una revista entendemos, según la clasificación de Surmanek⁵⁷, la totalidad de los lectores u hojeadores, bien sean primarios (los que viven en el hogar en que se compra la revista) o secundarios (su contrario), y bien la lean dentro o fuera del hogar.

La **selección de las revistas**, hasta 30, se realizó según la combinación de los criterios de cobertura (mayor audiencia o mayor Rating Point) y de rentabilidad (menor coste por mil impactos). Se combinaron los criterios sumando el escalafón ocupado por las revistas en cada criterio, de manera que se llegaba a una nueva clasificación en la que las revistas se ordenaban a la vez según la cobertura y según la rentabilidad. Las 30 revistas seleccionadas fueron, por este orden: *Teleprograma*, *Teleindiscreta*, *Pronto*, *Muy Interesante*, *Hola*, *Semana*, *TV Plus*, *Lecturas*, *El Jueves*, *Cruz Roja*, *Diez Minutos*, *Nuevo Vale*, *Clan TV*, *Interviú*, *Burda*, *Super Pop*, *Mía*, *Solomoto 30*, *Natura*, *Más Allá*, *Motociclismo*, *Autopista*, *Ragazza*, *Fotogramas*, *Tiempo*, *Cambio 16*, *Motor 16*, *Labores Hogar*, *Solo Moto Actual* y *Patrones*.

Como se puede ver, son de las anteriormente llamadas de información general; Agostini también estudiaba, en Francia, únicamente este tipo de revistas; más adelante Marc⁵⁸ analizó el comportamiento de las revistas especializadas francesas, descubriendo que el parámetro k variaba. Por ahora nuestro objetivo se centra sólo en validar la original fórmula de Agostini.

⁵⁷ SURMANEK, Jim. "Una guía fácil y rápida para la Planificación de Medios". Ed. Eresma. Madrid, 1982. Pág. 62.

De las 30 revistas, 19 tienen una periodicidad semanal, 1 quincenal y 10 mensual; el que una revista tire uno, dos o varios números al mes no va a influir en el análisis que nos ocupa. Del mismo modo, no se produce ningún sesgo por no seleccionar proporcionalmente las revistas entre las que nos ofrecía ODEC: 28 semanales, 5 quincenales y 64 mensuales. Sin más, se puede decir que la desigual proporción se debe a que se siguen otros criterios de selección y que la planificación de medios no pretende cuadrar en partes proporcionales.

El planteamiento de la investigación resulta original porque se estudia el comportamiento de los **públicos segmentados**, como hoy en día se dirigen las campañas publicitarias. Agostini se refería siempre a la población total, aunque en las conclusiones⁵⁹ apuntaba la conveniencia de analizar separadamente los "estratos" de la sociedad; años después, sería una grave omisión referirse únicamente a la población general. Buscando ser sistemático pero sin salirnos de la práctica corriente, se han considerado 19 públicos objetivo, el general y los 18 resultantes de combinar distintas alternativas de las variables sexo, edad y clase social. La configuración de los distintos segmentos, con el tamaño de la población y de la muestra, queda como refleja la **Tabla 2.3**:

⁵⁸ MARC, Marcel. "Étude de la Couverture Réalisée par Plusieurs Revues Techniques sur une Clientèle Spécialisée". Institut de Recherches et d'Études Publicitaires, estudio nº6. Noviembre 1962.

⁵⁹ AGOSTINI, Jean-Michel. "How Estimate Unduplicated Audiences". *Journal of Advertising Research*. Marzo 1961. Vol 1. Nº 3. Pág. 14.

Nº	SEXO	EDAD	CLASE SOCIAL	POBLACIÓN OBJETIVO	MUESTRA
1	hombre	14 - 34 años	alta / media alta	1.455.001	2.021
2	hombre	14 - 34 años	media media	3.067.620	4.257
3	hombre	14 - 34 años	media baja / baja	2.271.725	3.066
4	hombre	35 - 54 años	alta / media alta	1.065.349	1.127
5	hombre	35 - 54 años	media media	1.935.015	2.022
6	hombre	35 - 54 años	media baja / baja	1.562.292	1.602
7	hombre	≥ 55 años	alta / media alta	418.584	714
8	hombre	≥ 55 años	media media	1.226.080	2.084
9	hombre	≥ 55 años	media baja / baja	2.552.831	4.421
10	mujer	14 - 34 años	alta / media alta	1.299.247	1.767
11	mujer	14 - 34 años	media media	2.993.229	4.208
12	mujer	14 - 34 años	media baja / baja	2.252.025	2.822
13	mujer	35 - 54 años	alta / media alta	879.645	1.108
14	mujer	35 - 54 años	media media	1.897.628	2.438
15	mujer	35 - 54 años	media baja / baja	1.832.278	2.332
16	mujer	≥ 55 años	alta / media alta	518.613	739
17	mujer	≥ 55 años	media media	1.473.273	2.174
18	mujer	≥ 55 años	media baja / baja	3.299.573	4.958
19	total	total	total	32.000.008	43.860

Tabla 2.3: Públicos objetivo a los que se dirigen las campañas.

Los segmentos se han numerado como muestra la **Tabla 2.3**. Como se ve, la fila número 19 corresponde a la población entera, sin segmentar; por otra parte, esa fila es el sumatorio de las demás ya que los públicos objetivos son excluyentes y comprenden la totalidad de la población. Las variables que se han utilizado para segmentar son de carácter socio-demográfico (*sexo*, *edad* y *clase social*); se podrían haber considerado, con los datos del EGM, variables de consumo personal (dividen al público según consuman un producto o no), de equipamiento (posean o no un producto), etc; también otras socio-demográficas resultaban interesantes, como la de *rol familiar* (ama de casa, cabeza de familia y otros). Sin embargo se ha optado por las socio-demográficas porque su análisis permite una mejor generalización, al tiempo que no se desvirtúa la práctica corriente en la planificación de medios.

Para las variables *edad* y *clase social* se han elegido tres alternativas excluyentes que dividan la población total en subconjuntos lo suficientemente grandes como para que el análisis resulte representativo, contando con el tamaño muestral adecuado; como se observa en la última columna de la **Tabla 2.3**, en todos los casos resulta representativo.

Se prepararon 402 campañas, 20 por cada uno de los 18 targets y 42 en la población total. En la **Figura 2.2** las campañas se representan por cubos, definidos por tres dimensiones: el público objetivo, el número de inserciones y el criterio de selección de soportes.

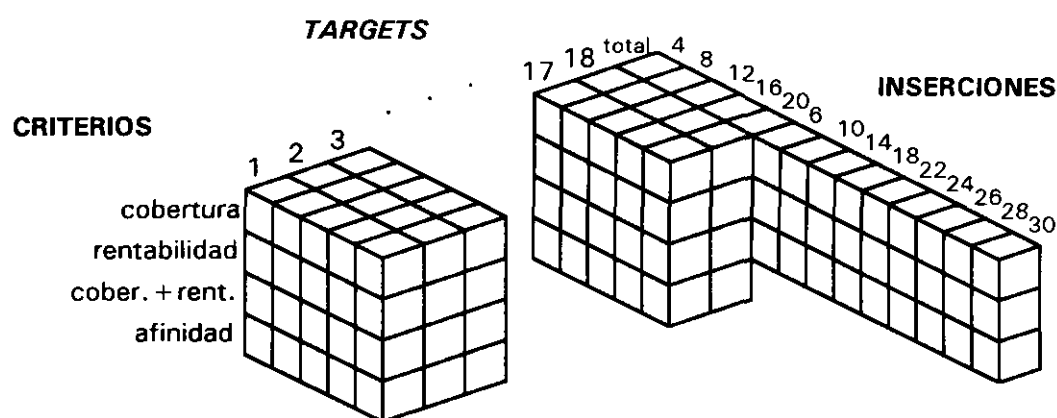


Figura 2.2

En la **selección de los soportes** de cada campaña se emplearon cuatro criterios: la cobertura, la rentabilidad, una combinación de ambos y la afinidad. Los primeros ya han sido comentados anteriormente (§C.1). El *coeficiente de afinidad* es el porcentaje de la audiencia total de un soporte que pertenece a nuestro público objetivo; veamos un ejemplo en la **Tabla 2.4**, para el *público 1* (hombres, de 14 a 34 años, de clases sociales alta y media alta), formado por 1.455.001 personas:

REVISTA	BURDA	SOLOMOTO 30
Audiencia en el <i>público 1</i>	10.900	92.500
Audiencia total	1.105.300	498.800
Coefficiente de afinidad	$\frac{10.900}{1.105.300} \approx 1\%$	$\frac{92.500}{498.800} \approx 18,5\%$

Tabla 2.4

Estamos ante dos casos extremos: la Revista BURDA tiene un coeficiente de afinidad del 1%, muy por debajo del porcentaje que el público objetivo ocupa sobre el total ($\frac{1.455.001}{32.000.000} = 4,5\%$), coeficiente que le correspondería si la audiencia se distribuyese de manera homogénea. Al contrario, SOLOMOTO 30 cuenta con un 18,5% de sus lectores en el público 1, muy por encima de la media de otros públicos. Dejando ya el ejemplo, el criterio de afinidad consiste en seleccionar aquellos soportes que tengan ese coeficiente más alto. Sobre la población total, el coeficiente valdrá el 100% en todos los soportes: luego no se podrá emplear este criterio.

Para seleccionar los soportes se pidió a ODEC la clasificación de las 30 revistas (TOM1) para cada uno de los 19 *targets*. Público a público, se escogían las 4 revistas con mayor cobertura, rentabilidad, cober.+rent. y afinidad, las 8 revistas con mayor cobertura, rentabilidad, cober.+rent. y afinidad, y así sucesivamente. Todas las campañas, en consecuencia, se han planificado según criterios racionales, sin que intervengan variables aleatorias. Se trata de una diferencia con respecto a Agostini, pues él determinaba las combinaciones de soportes al azar. Nos parece más adecuado utilizar criterios racionales pues así es la práctica corriente de las Agencias.

Como planteaban otras investigaciones⁶⁰, se ha considerado distinto número de inserciones en las campañas, para estudiar la repercusión que pueda tener esta variable.

⁶⁰ BOWER, John. "Net Audiences of U.S..." (op.cit.); CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British..." (op.cit.); CLAYCAMP, H.J. and McCLELLAND, C.W. "Estimating Reach..." (op.cit.); MARC, Marcel. "Net Audiences of French..." (op.cit.).

En concreto, principalmente hay cinco grupos con 4,8,12,16 y 20 inserciones, cada una con 75 casos; el total se completa con 27 campañas, con 6, 10, 14, 18, 22, 24, 26, 28 y 30 inserciones (**Figura 2.2**). Este diseño presenta una deficiencia que he confirmado cuando el trabajo ya estaba avanzado: al seguir estrictamente los criterios, algunas combinaciones seleccionadas según la cober.+rent. coincidían con las seleccionadas por los criterios de cobertura o de rentabilidad. Ello provoca que las combinaciones de soportes realmente distintas sean unas 370. No obstante, los casos repetidos no son inútiles, porque permiten analizar las diferencias entre criterios.

Las **duplicaciones** se han tomado del programa **TOM1D** (Duplicación entre Soportes). Se obtuvieron 19 salidas de ordenador, una por público objetivo; en cada salida se recogían las 435 duplicaciones posibles de las treinta revistas seleccionadas.

La cantidad de operaciones realizadas ha sido considerable: cada una de las 402 campañas tiene un sumatorio de duplicaciones con 100 términos de media; el trabajo resulta más completo que el de Agostini, gracias a que hoy en día contamos con muchos más medios informáticos. Se ha utilizado la hoja de cálculo Excel 3.0.

Los **alcances brutos y netos** se obtuvieron mediante el programa **TOM2** (Evaluación de Planes). Se evaluaron las combinaciones de soportes previstas -en nuestro caso con una sola inserción por soporte-. La salidas nos ofrecen el alcance neto o número de impactados, es decir, personas distintas expuestas independientemente de cuantas veces, y el alcance bruto o número de impactos o, dicho de otra forma, las veces que se ha contactado independientemente de qué personas.

1.4. RESULTADOS DE LA INVESTIGACIÓN

1.4.1. Aplicación de la fórmula de Agostini

Tras comentar la metodología de Agostini, se aplica el mismo proceso a nuestros datos; los resultados se recogen en el fichero **ALCANCE.WK1** del diskette anejo. En las primeras columnas se numeran las campañas y se recogen sus características (público objetivo, criterio de selección de soportes y nº de inserciones); después aparece el alcance bruto (antes llamado A), la suma de duplicaciones (D) y el alcance neto estimado por TOM2 (C).

Agostini había encontrado una relación entre x y Z , tal que

$$Z = \frac{1}{kx + 1} \quad \text{siendo,} \quad x = \frac{D}{A} \quad \text{y} \quad Z = \frac{C}{A}$$

Calculando x y Z a partir de nuestros datos, se obtiene una relación similar. Al dibujar sus valores sobre dos ejes, se observa que la nube de los 402 puntos se puede ajustar mediante una hipérbola equilátera, asintótica con el eje x (cfr. **Figura 2.3**).

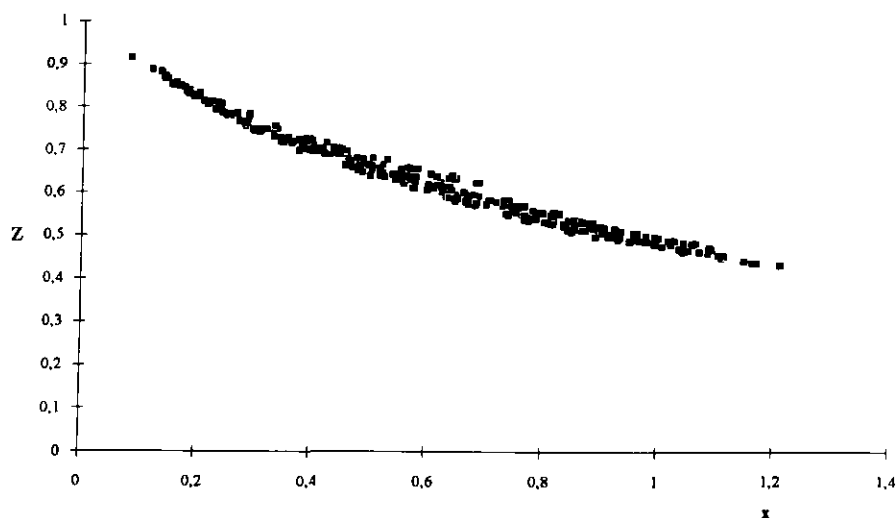


Figura 2.3

Primeramente hemos transformado la anterior relación entre x y Z, en un modelo lineal como sigue:

$$\frac{1}{Z} = kx + 1$$

Hemos realizado la regresión con el SPSS/PC+, siendo la 1/Z la variable explicada y x la variable explicativa. Los resultados se recogen en la **Figura 2.4**:

* * * * M U L T I P L E R E G R E S S I O N * * * *					
Multiple R		0,99228			
R Square		0,98463			
Adjusted R Square		0,98459			
Standard Error		0,03493			
Analysis of Variance					
	DF	Sum of Squares	Mean Square	Regression	
Regression	1	31,25661	31,25661		
Residual	400	0,48802	0,00122		
F =	25618,98962	Signif F =	0,0000		
Variable	B	SE B	Beta	T	Sig T
H	1,048378	0,006550	0,992284	160,059	0,0000
(Constant)	0,995700	0,004446		223,961	0,0000

Figura 2.4

La nueva variable dependiente es 1/Z (de fácil cálculo) será explicada por la variable independiente x, en un modelo como sigue:

$$\frac{1}{Z} = \alpha + \beta \cdot x$$

Los coeficientes α y β determinan la relación lineal entre las variables, minimizando los errores del ajuste. Sustituyendo por los resultados obtenidos, el modelo quedaría:

$$\frac{1}{Z} = 0,9957 + 1,048378 \cdot x$$

Los resultados de la regresión son muy satisfactorios; el R^2 mayor del 98% indica que el ajuste explica muy bien la dispersión de los datos. Junto a esto, con todos los estadísticos significativos, se puede afirmar que la hipérbola equilátera ajusta adecuadamente la nube de puntos. Aceptando el pequeño margen de error, se puede decir que el parámetro k (antes llamado β) toma el valor 1'048.

Una vez determinado $k=1'048$, vamos a seguir el proceso inverso. A partir de los alcances brutos y las sumas de duplicaciones podemos estimar el alcance neto. Posteriormente hemos analizado los errores de la estimación, comparándolos con el alcance neto observado en el TOM2. Habíamos concluido (**I Parte, §D**) que el TOM era el modelo que mejor estimaba el alcance, y particularmente mejor en este tipo de campañas. El error se mide, ahora y en sucesivos modelos, como sigue:

$$Error (\%) = \frac{\text{alcance estimado} - \text{alcance observado en TOM 2}}{\text{alcance observado en TOM 2}} \cdot 100$$

Difiere con la medida empleada por Bower⁶¹, pero ésta nos parece más adecuada, por ser más precisa y frecuente en la literatura. En las siguientes tablas aparecen los errores de estimación, cruzados con las distintas variables que hemos empleado en nuestro estudio. En los errores hay que analizar tanto la magnitud como el signo: de ahí que se recojan separadamente los errores en valor absoluto. El alcance neto estimado por Agostini ($k=1'048$) se recoge en el fichero **ALCANCE.WK1**.

Comenzamos con el análisis de los errores del modelo de Agostini (con $k=1,048$), cruzados con los distintos públicos objetivos (**Tabla 2.5**). Primeramente se observa que la media de todos los errores es del -0,23%, es decir, el modelo produce subestimaciones del alcance neto, pero de muy poca magnitud; la desviación típica del 2% nos indica que los errores habitualmente se situarán dos puntos porcentuales por encima o por debajo de la media. Pero la media ha anulado los signos opuestos de los datos y, por eso, hemos

⁶¹ BOWER, John. "Net Audiences of U.S. and Canadian Magazines: Seven Test of Agostini's Formula". *Journal of Advertising Research*. Marzo, 1963. Vol 3. Nº 1. Págs. 13-20.

de fijarnos en la media de los errores en valor absoluto: el valor de 1,64% señala el error medio de todas las observaciones, independientemente de si son sobrestimaciones o subestimaciones. Al comparar estas magnitudes en cada público objetivo observamos que no presentan unas desviaciones sistemáticas, lo que nos da a entender que el modelo estima indistintamente según el *target* al que se dirija la campaña.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	1,1850	0,8025	1,2815	0,6274
H, 14-34, MM	1,0195	1,0647	1,2283	0,7997
H, 14-34, MB-B	0,6873	1,0188	1,0587	0,5961
H, 35-54, A-MA	-0,7891	1,1549	1,1129	0,8281
H, 35-54, MM	-0,5305	1,2225	1,1081	0,7058
H, 35-54, MB-B	-2,0138	2,2010	2,2646	1,9273
H, >54, A-MA	-1,6111	1,1055	1,6761	0,9986
H, >54, MM	-4,2199	2,4729	4,2288	2,4568
H, >54, MB-B	-2,1854	1,4719	2,2540	1,3586
M, 14-34, A-MA	1,4555	0,7447	1,4683	0,7177
M, 14-34, MM	1,5943	0,7709	1,6189	0,7150
M, 14-34, MB-B	1,6002	0,5925	1,6002	0,5925
M, 35-54, A-MA	-0,2838	1,1604	1,1097	0,3673
M, 35-54, MM	1,0750	0,5929	1,1563	0,4007
M, 35-54, MB-B	-1,3323	1,6575	1,8309	1,0427
M, >54, A-MA	2,4117	1,2304	2,4117	1,2304
M, >54, MM	0,8048	0,7209	0,8961	0,5970
M, >54, MB-B	-1,2881	1,2312	1,6605	0,6000
TOTAL	-1,1258	1,1344	1,4673	0,6160
TODOS LOS CASOS	-0,2381	2,0417	1,6441	1,2311

Tabla 2.5: Errores del modelo de Agostini ($k=1,048$), según el público objetivo.

Del mismo modo vemos, en la **Tabla 2.6**, que tampoco parece existir relación entre los errores y el criterio de selección de soportes. No aparecen variaciones sistemáticas con respecto a la media, ni en cuanto a la magnitud, signo ni dispersión de los errores.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-0,0771	1,9665	1,5755	1,1691
cobertura	-0,3493	2,1694	1,8108	1,2324
rentabilidad+cobertura	-0,1577	2,1045	1,7267	1,2016
afinidad	-0,3882	1,9130	1,4354	1,3147
TODOS LOS CASOS	-0,2381	2,0417	1,6441	1,2311

Tabla 2.6: Errores del modelo de Agostini ($k=1,048$), según el criterio de selección de soportes.

Numerosos investigadores habían señalado que el modelo de Agostini debía aplicarse de manera distinta según el número de soportes. Lo mismo se deduce de nuestra investigación. Como muestra la **Tabla 2.7**, el error se hace cada vez mayor cuando crece el número de soportes de la campaña: aproximadamente desde el 1% con 4 soportes hasta el 2,2% con 20 soportes; con pocas revistas se producen sobrestimaciones, y gradualmente con muchas revistas hay subestimaciones. También la desviación estándar de los errores crece con el número de soportes, de modo que los resultados están más dispersos. En la **Tabla 2.7**, no se recogen 27 campañas, formadas por 6, 10, 14, 18, 22, 24, 26, 28 ó 30 revistas; se ha hecho así, para evitar sesgos y no hacer demasiado engorrosa la presentación de los datos. No obstante, esas campañas mantienen las tendencias aquí apuntadas.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	0,8772	0,8023	1,0181	0,6107
8	0,1786	1,4555	1,1999	0,8316
12	-0,2742	2,0792	1,7714	1,1040
16	-0,5906	2,3545	1,9835	1,3819
20	-0,9580	2,6498	2,2643	1,6599
TODOS LOS CASOS	-0,2381	2,0417	1,6441	1,2311

Tabla 2.7: Errores del modelo de Agostini ($k=1,048$), según el número de soportes.

Hemos encontrado el valor del parámetro k para España, y se han analizado los errores de estimación que produce. Pero ¿qué errores se habrían cometido si hubiésemos tomado $k=1,125$ como en otros países? Se habrían obtenido subestimaciones en casi todos los casos, y la media de los errores absolutos sería del 2,9%, con una desviación estándar del 2,1%. Por tanto, los errores casi se habrían duplicado.

1.4.2. Análisis a posteriori del modelo de Agostini

También podemos determinar el valor de k **a posteriori**, es decir, despejándolo en cada caso pues conocemos todos los términos de la ecuación. La media de k así determinada ha sido 1,044. Pero no es del todo correcto hablar de un parámetro general: como habíamos intuido el parámetro no se comporta de manera constante: varía, por una parte, según la población objetivo y, por otra, según el número de inserciones, mientras que es constante en los distintos criterios utilizados. Mediante el SPSS hemos cruzado los valores de k con las otras variables estudiadas.

En la **Tabla 2.8**, se recoge la media del parámetro k para cada público objetivo, dando la impresión de que existen diferencias relevantes que luego se intentarán probar como estadísticamente significativas.

POBLACIÓN OBJETIVO	MEDIA K	DESVIACIÓN TÍPICA	Nº DE CAMPAÑAS
H, 14-34, A-MA	1,0820	0,0246	20
H, 14-34, MM	1,0782	0,0371	20
H, 14-34, MB-B	1,0713	0,0383	20
H, 35-54, A-MA	1,0356	0,0387	20
H, 35-54, MM	1,0411	0,0398	20
H, 35-54, MB-B	0,9944	0,0640	20
H, >54, A-MA	1,0054	0,0372	20
H, >54, MM	0,9321	0,0590	20
H, >54, MB-B	0,9745	0,0507	20
M, 14-34, A-MA	1,0873	0,0222	20
M, 14-34, MM	1,0897	0,0245	20
M, 14-34, MB-B	1,0918	0,0227	20
M, 35-54, A-MA	1,0490	0,0352	20
M, 35-54, MM	1,0770	0,0235	20
M, 35-54, MB-B	1,0198	0,0486	20
M, >54, A-MA	1,1072	0,0219	20
M, >54, MM	1,0731	0,0271	20
M, >54, MB-B	1,0168	0,0430	20
TOTAL	1,0273	0,0321	42
TODOS LOS CASOS	1,0440	0,0576	402

Tabla 2.8: Medias del parámetro k según cada público.

Por otra parte, parece que el criterio de selección de soportes no ha influido en el valor del parámetro k (cfr. **Tabla 2.9**). Todos los criterios empleados se acercan bastante al valor medio de k para todos los casos.

CRITERIO	MEDIA K	DESVIACIÓN TÍPICA	Nº DE CASOS
rentabilidad	1,0485	0,0551	104
cobertura	1,0421	0,0617	104
rentabilidad+cobertura	1,0468	0,0584	104
afinidad	1,0376	0,0547	90
TODOS LOS CASOS	1,0440	0,0576	402

Tabla 2.9: Medias del parámetro k según el criterio de selección de soportes.

Del mismo modo se ha estudiado la influencia del número de revistas sobre el parámetro k (**Tabla 2.10**), observándose que están relacionados negativamente: cuantos

más soportes tenga la campaña, menor será el valor de k . Coincide con las conclusiones obtenidas por Caffyn y Sagovsky para los periódicos ingleses⁶².

Nº SOPORTES	MEDIA K	DESVIACIÓN TÍPICA	Nº DE CASOS
4	1,0906	0,0384	75
6	1,0720	0,0181	3
8	1,0508	0,0463	75
10	1,0415	0,0241	3
12	1,0368	0,0567	75
14	1,0075	0,0054	3
16	1,0293	0,0594	75
18	1,0129	0,0063	3
20	1,0211	0,0643	75
22	1,0127	0,0045	3
24	1,0117	0,0025	3
26	1,0094	0,0008	3
28	1,0027	0,0085	3
30	1,0048	0,0003	3
TODOS LOS CASOS	1,0440	0,0576	402

Tabla 2.10: Media del parámetro k según el número de soportes.

Pero desde un punto de vista metodológico no es suficiente con visualizar las diferentes medias que se muestran en las tablas. Si queremos llegar a conclusiones generalizables hemos de utilizar instrumentos adecuados de análisis. Para ello, se ha empleado la técnica ANOVA⁶³, con el paquete estadístico SPSS/PC+. Se trataba de analizar la varianza para conocer la influencia de las variables *población objetivo*, *criterio* y *nº de inserciones* en el parámetro k : como difieren las medias según cada variable, hay que investigar si las diferencias se deben a variaciones aleatorias o si existen diferencias sistemáticas. Sin pretender explicar aquí la técnica estadística, pues no es materia de nuestro estudio, se explicarán brevemente los términos empleados.

⁶² CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods". *Journal of Advertising Research*. Marzo 1963. Vol 3. Nº 1. Págs. 21-25.

⁶³ INVERSEN, G.R. y NORPORTH, H. "Analysis of Variance". Quantitative Applications in the Social Sciences Nº 1. A Sage University Paper Series.

La *Variación Entre* recoge la variación de la media de cada uno de los grupos respecto a la media total, teniendo $n-1$ grados de libertad ($n = \text{nº de grupos}$). Se calcula de la siguiente manera:

$$V.E. = \sum_{j=1}^n p_j (\bar{k}_j - \bar{k})^2$$

donde,

\bar{k}_j = media aritmética del grupo j .

p_j = ponderación del número de elementos de cada grupo j .

\bar{k} = media aritmética total del parámetro k .

$j = 1, 2, \dots, n$ grupos.

La *Variación Intra* es la diferencia entre una observación y la media del grupo a la que pertenece. También es llamada Variación Intragrupo, ya que mide la homogeneidad de cada uno de los grupos. Presenta tantos grados de libertad como observaciones menos nº de grupos: $m-n$. Matemáticamente se define así:

$$V.I. = \sum_{j=1}^n \sum_{i=1}^m (k_{ij} - \bar{k}_j)^2$$

donde,

k_{ij} = valor que toma k en la observación i del grupo j

$i = 1, 2, \dots, m$ observaciones.

La *Variación Total* es la suma de los dos componentes anteriores. Mide las diferencias que presentan las observaciones respecto a la media aritmética total. Tendrá $m-1$ grados de libertad, siendo m el nº de observaciones. Su formulación quedaría como sigue:

$$V.T. = \sum_{j=1}^n \sum_{i=1}^m (k_{ij} - \bar{k})^2$$

Si la Variación Intra es igual a la Variación Total, no existirán diferencias entre los grupos. Si las diferencias entre cada grupo son nulas, el ratio F estimado estaría cerca de la unidad. Por el contrario, si las diferencias son no nulas, F tomará valores superiores a la unidad. Se realizará un contraste de hipótesis, donde la hipótesis nula (H_0) es que las diferencias son nulas y la hipótesis alternativa (H_1) que alguna(s) de ellas no es(son) nula(s).

Dado que la F se distribuye como una F de Snedecor con $(n-1, m-n)$ grados de libertad, podemos efectuar la comparación entre el ratio calculado y el valor de las tablas, concluyendo que :

si $F > F_e \Rightarrow H_1$ (existen diferencias entre los grupos)

si $F \leq F_e \Rightarrow H_0$ (no existen diferencias entre los grupos)

Las salidas del SPSS/PC+ se han ordenado en las siguientes tablas, para que resulte más fácil su visualización. Son cuatro tablas, donde la variable que queremos explicar es el parámetro k y las variables explicativas la *población objetivo*, el *criterio de selección de soportes* y el *nº de inserciones*.

	suma de cuadrados	grados de libertad	media al cuadrado	F	signific. de F	F en tablas
<i>Var. Entre</i>	0,777	18	0,043	29,913	0,000	1,642
<i>Var. Intra</i>	0,553	383	0,001			
<i>Var. Total</i>	1,330	401	0,003			

Tabla 2.11: ANOVA: Parámetro k respecto al público objetivo.

Como se puede observar en la **Tabla 2.11**, el valor estimado de F es mayor que la F de tablas ($29,913 > 1,642$); luego se acepta la hipótesis alternativa, es decir, que existen diferencias entre cada uno de los grupos: el parámetro k se comporta de manera

distinta dependiendo de la población objetivo; esto lleva inmediatamente a pensar que no se puede hablar de un parámetro general para un medio y un país (tesis de Agostini), pero más adelante se matizará esta cuestión.

	suma de cuadrados	grados de libertad	media al cuadrado	F	signific. de F	F en tablas
Var. Entre	0,007	3	0,002	0,705	0,549	2,632
Var. Intra	1,323	398	0,003			
Var. Total	1,330	401	0,003			

Tabla 2.12: ANOVA: Parámetro k respecto al criterio de selección de soportes.

Al analizar la influencia del factor criterio de selección de soportes (**Tabla 2.12**), nos encontramos con una clara diferencia respecto al caso anterior: ahora $F < F_{\epsilon}$ ($0,705 < 2,632$), luego se cumple la hipótesis nula H_0 de que no existen diferencias entre las distintas alternativas de la variable *criterio de selección de soportes*; es decir, resulta indiferente para el parámetro k que la planificación se haya realizado buscando la mayor rentabilidad, la mayor cobertura o la mayor afinidad. Se demuestra lo que antes se intuía al ver las pequeñas diferencias en las medias de los diferentes grupos. También se puede observar que la pequeña dimensión de la Variación Entre (0,007) indica pocas diferencias entre los grupos, frente a una Variación Intra (1,323) que explica que casi toda la Variación Total está determinada por las diferencias dentro de cada grupo.

	suma de cuadrados	grados de libertad	media al cuadrado	F	signific. de F	F en tablas
Var. Entre	0,255	13	0,020	7,068	0,000	1,767
Var. Intra	1,075	383	0,003			
Var. Total	1,330	401	0,003			

Tabla 2.13: ANOVA: Parámetro k respecto al número de inserciones (I).

El ANOVA del parámetro k respecto al número de inserciones (I) nos muestra, en la **Tabla 2.13**, que se vuelve a cumplir la hipótesis alternativa H_1 ya que $F > F_{\epsilon}$ ($7,068 > 1,767$), de manera que los catorce grupos considerados presentan diferencias

significativas sobre la media total. Pero queremos evitar el problema de muchos de esos grupos no tienen un número representativo de casos: por eso, se realizó el mismo análisis pero con sólo 5 grupos, cada uno con 75 casos (cfr. **Tabla 2.14**).

	suma de cuadrados	grados de libertad	media al cuadrado	F	signific. de F	F en tablas
<i>Var. Entre</i>	0,225	4	0,056	19,385	0,000	2,403
<i>Var. Intra</i>	1,073	370	0,003			
<i>Var. Total</i>	1,298	374	0,003			

Tabla 2.14: ANOVA: Parámetro k respecto al número de soportes (II).

La **Tabla 2.14** nos ratifica las conclusiones de la anterior, reforzándose incluso ya que la nueva F es considerablemente mayor a la F en tablas. Por tanto se vuelve a demostrar que habría que considerar un parámetro k distinto según el número de revistas donde se ha insertado el anuncio. Llegamos aquí a la misma conclusión apuntada por otros autores.

1.4.3. Aplicación con el parámetro k variable

Tras demostrar estadísticamente que k no es constante, nos planteamos primero que varíe *en función del público objetivo*. En vez de tomar el valor $k=1'048$, se va a tomar el que corresponda a la media de cada público, recogido en la segunda columna de la **Tabla 2.8**. Estos valores se han determinado *a posteriori*, despejando el parámetro k de la fórmula de Agostini, conocido ya el alcance neto observado. Esto es más adecuado que plantear para cada *target* la regresión hecha por Agostini, porque 20 son pocos casos para ajustar una curva.

Al estimar la audiencia neta con esos valores de k , hemos obtenido los errores sobre el alcance neto observado. Esos resultados se recogen en las tres tablas sucesivas. Hacer que el parámetro k variase según el *target*, ha mejorado el modelo en comparación con un parámetro constante: antes la media de errores absolutos era del 1,6%, con una desviación típica del 1,2%, y ahora es del 0,9%, con una desviación del 6%. Aproximadamente los errores se han reducido a la mitad. El signo de estos errores es negativo, luego tiende a la subestimación. La **Tabla 2.15** recoge los errores de estimación para cada público objetivo: son distintos pero su variación es relativamente pequeña; sólo en cinco casos la media de errores es algo superior al 1%. No encontramos relación entre las variables constitutivas del *target* y los errores.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	-0,0624	0,7778	0,5814	0,5032
H, 14-34, MM	-0,0899	1,0924	0,8909	0,6053
H, 14-34, MB-B	-0,1228	1,0696	0,8815	0,5849
H, 35-54, A-MA	-0,3700	1,0399	0,8964	0,6162
H, 35-54, MM	-0,3034	1,1714	1,0275	0,5981
H, 35-54, MB-B	-0,4612	1,8572	1,5744	1,0315
H, >54, A-MA	-0,2706	0,7948	0,6021	0,5720
H, >54, MM	-0,5378	1,5167	1,4562	0,6105
H, >54, MB-B	-0,3484	0,9752	0,9665	0,3091
M, 14-34, A-MA	-0,0277	0,6338	0,4924	0,3837
M, 14-34, MM	-0,0721	0,7784	0,6006	0,4813
M, 14-34, MB-B	-0,1790	0,8578	0,7078	0,4921
M, 35-54, A-MA	-0,3231	1,1687	1,1259	0,3760
M, 35-54, MM	-0,0616	0,6187	0,4384	0,4295
M, 35-54, MB-B	-0,3384	1,4417	1,3034	0,6409
M, >54, A-MA	0,1158	0,8179	0,7481	0,3066
M, >54, MM	-0,1793	0,8712	0,6896	0,5406
M, >54, MB-B	-0,2723	1,0352	0,8913	0,5599
TOTAL	-0,2756	0,9588	0,9012	0,4064
TODOS LOS CASOS	-0,2230	1,0507	0,8839	0,6088

Tabla 2.15: Errores del modelo de Agostini (k variable en función del *target*), según el público objetivo.

En relación con el criterio de selección de soportes, se observa que los errores también se han reducido aproximadamente a la mitad, en la rentabilidad, la cobertura, la combinación de ambos y la afinidad (cfr. **Tabla 2.16**). Los mayores errores corresponden al criterio de la cobertura y los menores al de la rentabilidad. Hay que decir, no obstante, que tratándose de una diferencia tan pequeña, podemos obviar el efecto de criterio.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-0,0399	0,9342	0,7586	0,5415
cobertura	-0,3071	1,1399	0,9931	0,6315
rentabilidad+cobertura	-0,1224	1,0623	0,8709	0,6146
afinidad	-0,4537	1,0206	0,9176	0,6312
TODOS LOS CASOS	-0,2230	1,0507	0,8839	0,6088

Tabla 2.16: Errores del modelo de Agostini (k variable en función del *target*), según el criterio de selección de soportes.

Muy significativo resulta el análisis del cruce entre los errores y el número de soportes de la campaña. Recordamos que, con la constante $k=1'048$, los errores aumentaban al emplear más revistas. Ahora, variando k en función del público objetivo, los errores no siguen ninguna tendencia, permaneciendo bastante estables (cfr. **Tabla 2.17**). Estamos, por tanto, ante un método que no está afectado por el número de soportes que compone la campaña: en comparación con otros métodos, se trata de una indudable ventaja. También se advierte que la mejor estimación de este método viene de la corrección de los errores cuando se empleaban muchos soportes. Para 4 y 8 revistas los errores eran similares al de contante $k=1'048$ (véase **Tabla 2.7**), mientras que para 16 y 20 revistas los errores en valor absoluto se han reducido a la mitad.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	0,8315	0,7957	1,0214	0,5258
8	0,1508	0,9321	0,7577	0,5569
12	-0,3105	0,7158	0,5917	0,5052
16	-0,6496	0,7962	0,8558	0,5651
20	-1,0347	0,9598	1,1791	0,7728
TODOS LOS CASOS	-0,2230	1,0507	0,8839	0,6088

Tabla 2.17: Errores del modelo de Agostini (k variable en función del *target*), según el número de soportes.

Planteamos a continuación los resultados de la estimación de la fórmula de Agostini con un k que varía *en función del número de soportes*. Para obtener los valores de k , hemos realizado la misma regresión que se muestra en la **Figura 2.4**, pero

separadamente para 4, 8, 12, 16 y 20 revistas (tenemos 75 casos en cada alternativa).

Recordamos que se realiza una regresión simple con la siguiente forma:

$$\frac{1}{Z} = \alpha + \beta \cdot x \quad \text{donde} \quad Z = \frac{C}{A} \quad \text{y} \quad x = \frac{D}{A}$$

En la **Tabla 2.18** se recogen los parámetros α y β ($\beta=k$) que ajustan la nube de puntos correspondientes, así como la significación del modelo y de los parámetros:

Número de soportes	Parámetro k	Constante α	R ² ajustado	Significación del modelo	Significación de los parámetros
4	1,1424	0,9883	0,9893	Significativo	Significativos
8	1,1392	0,9589	0,9733	Significativo	Significativos
12	1,2098	0,8914	0,9654	Significativo	Significativos
16	1,2136	0,8620	0,9714	Significativo	Significativos
20	1,2255	0,8300	0,9762	Significativo	Significativos

Tabla 2.18

Se puede advertir que el parámetro k se hace mayor al aumentar el número de soportes. Ahora bien, el análisis *a posteriori* (cfr. **Tabla 2.10**) nos mostraba la relación contraria, con unos valores próximos a la unidad. Además estos valores de k divergen bastante de la constante $k=1,048$ que ha servido para estimar adecuadamente la audiencia neta, mejor que el valor empleado en otros países de $k=1,125$. ¿Por qué entonces las regresiones según el número de soportes nos ha dado valores de k tan distintos? Porque la constante de la recta de regresión ha variado de modo relevante, como se observa en la tercera columna de la **Tabla 2.18**. Para Agostini esa constante debía ser la unidad. Modificando ligeramente la fórmula de Agostini, podremos mejorar las estimaciones. Veámoslo para cuatro soportes:

$$\frac{1}{Z} = 0,9884 + 1,1424 \cdot x$$

Sustituyendo Z y x , obtenemos:

$$\frac{A}{C} = 0,9884 + 1,1424 \cdot \frac{D}{A} \Leftrightarrow C = \frac{A}{0,9884 + 1,1424 \cdot \frac{D}{A}}$$

Del mismo modo, la fórmula de Agostini queda definida así según el número de soportes:

$$\text{Para 8 soportes: } C = \frac{A}{0,9589 + 1,1392 \cdot \frac{D}{A}}$$

$$\text{Para 12 soportes: } C = \frac{A}{0,8914 + 1,2098 \cdot \frac{D}{A}}$$

$$\text{Para 16 soportes: } C = \frac{A}{0,862 + 1,2136 \cdot \frac{D}{A}}$$

$$\text{Para 20 soportes: } C = \frac{A}{0,83 + 1,2255 \cdot \frac{D}{A}}$$

A partir de estas fórmulas hemos estimado el alcance neto de las campañas; ahora sólo se consideran 375 casos porque el resto tiene distinto número de soportes. Los errores de estimación se recogen en las tablas siguientes. Se puede ver como las sobrestimaciones se contrapesan con las subestimaciones, de manera que la media de los errores es casi nula; sin embargo la media de los errores absolutos (1,1%) muestra la dimensión de los errores, independientemente de su signo. La dispersión respecto a la media de esos errores es del $\pm 1\%$. Esta modificación del modelo de Agostini ha supuesto una mejora (antes la media de errores absolutos era del 1,6%), pero no es tan buena como la estimación con k variable en función del *target* (0,9%).

En la **Tabla 2.19** recogemos las medias y desviaciones típicas sobre los públicos objetivo. Se observan diferencias que no tienen explicación en función de el sexo, la edad o la clase social. Pensamos que con estimaciones tan precisas no es necesario acentuar el énfasis en estas pequeñas diferencias. Como nos sucederá en otros modelos, el público formado por los hombres, mayores de 54 años, de clase social media media, se comporta de forma atípica, sin que se haya podido alcanzar una explicación racional.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
H, 14-34, A-MA	0,8659	0,8699	1,0103	0,6868
H, 14-34, MM	0,7862	1,1422	1,1883	0,6846
H, 14-34, MB-B	0,9249	1,0027	1,1785	0,6664
H, 35-54, A-MA	-0,3948	1,1333	0,9154	0,7526
H, 35-54, MM	0,1638	0,9018	0,7125	0,5543
H, 35-54, MB-B	-0,2471	1,1468	0,9288	0,6866
H, >54, A-MA	-0,5509	1,2399	1,0589	0,8225
H, >54, MM	-3,2375	1,4803	3,2375	1,4803
H, >54, MB-B	0,6631	1,3506	1,1581	0,9356
M, 14-34, A-MA	0,7824	1,0992	1,0516	0,8298
M, 14-34, MM	0,4931	0,9865	0,8478	0,6874
M, 14-34, MB-B	0,4387	0,7933	0,7046	0,5562
M, 35-54, A-MA	-1,1760	1,4290	1,6134	0,8706
M, 35-54, MM	0,3044	0,6015	0,5756	0,3326
M, 35-54, MB-B	-1,1861	1,0780	1,3667	0,8226
M, >54, A-MA	1,9214	1,8437	2,1166	1,6028
M, >54, MM	0,2351	0,4123	0,3928	0,2563
M, >54, MB-B	-0,2091	0,9478	0,7253	0,6249
TOTAL	-0,7110	0,9127	0,9830	0,5826
TODOS LOS CASOS	0,0024	1,5358	1,1477	1,0188

Tabla 2.19: Errores del modelo de Agostini (k variable según el número de soportes), según el público objetivo.

Los errores de estimación presentan una mínima variación según se haya empleado uno u otro criterio de selección de soportes (véase la **Tabla 2.20**). Cuando se busca la rentabilidad y la afinidad, se han obtenido ligeras sobrestimaciones, mientras que al buscar la cobertura, la media obtenida muestra pequeñas subestimaciones.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	0,3013	1,3956	1,0742	0,9345
cobertura	-0,3763	1,6239	1,2537	1,0917
rentabilidad+cobertura	-0,0484	1,5155	1,1098	1,0268
afinidad	0,1401	1,5405	1,1535	1,0235
TODOS LOS CASOS	0,0024	1,5358	1,1477	1,0188

Tabla 2.20: Errores del modelo de Agostini (k variable en función del número de soportes), según el criterio de selección de soportes.

La **Tabla 2.21** muestra la relación positiva entre el número de soportes y los errores. Cuando la campaña recoge más revistas, el tamaño y la dispersión de los errores aumenta. En todas las alternativas las sobrestimaciones se equilibran con las subestimaciones, como se puede ver en la segunda columna de la tabla. Es relevante que no se haya quitado la tendencia creciente de los errores, usando un k variable en función del número de soportes; sin embargo, ese efecto se consiguió usando un k variable en función del *target*. Ello nos refuerza en que el método anterior es el que mejor estima el alcance neto.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	-0,0003	0,6454	0,5070	0,3950
8	0,0026	1,3225	1,0390	0,8092
12	0,0025	1,6793	1,2879	1,0672
16	0,0006	1,7856	1,4104	1,0828
20	0,0064	1,9357	1,4944	1,2180
TODOS LOS CASOS	0,0024	1,5358	1,1477	1,0188

Tabla 2.21: Errores del modelo de Agostini (k variable en función del número de soportes), según el número de soportes.

2. La fórmula de Kuhn

2.1. Desarrollo del autor

El investigador alemán encontró, tras estudiar varias revistas de su país, que la ecuación para estimar el alcance neto de una combinación de soportes, depende del número de soportes empleado⁶⁴. Primeramente aplicó la fórmula de Agostini a los datos tomados del Arbeitsgemeinschaft Leseranalyse, obteniendo una media de errores de más del 5%, mucho mayor que los obtenidos en Francia con el CESP. Dándole a k el valor de 1,162, se conseguía reducir el error hasta algo más del 2%. Encontró que la hipérbola que relacionaba x y z producía más errores que una función exponencial de esta forma:

$$z = e^{-k_n x}$$

que gráficamente se representa como aparece en la **Figura 2.5**:

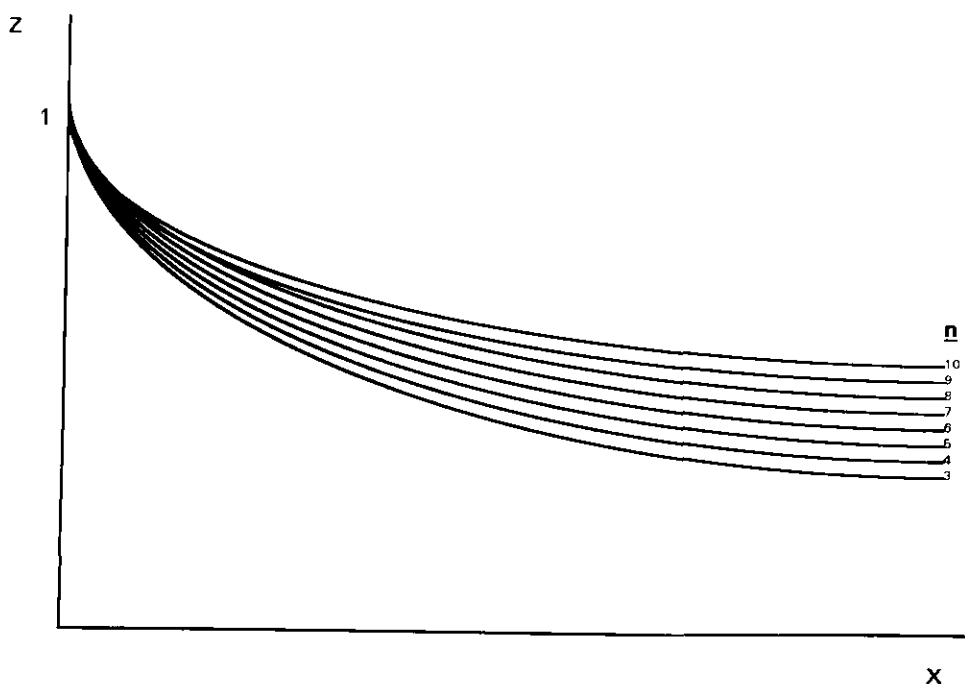


Figura 2.5

⁶⁴ KUHN, Walter. "Net Audiences of German Magazines: A New Formula". *Journal of Advertising Research*. Marzo 1963. Vol. 3. Nº 1. Págs. 30-33.

Los coeficientes k_n variaban según el número n de soportes, como se muestra en la **Figura 2.5**. Sus valores se obtuvieron empíricamente, y mostraban una secuencia que sugería el ajuste a una fórmula general, como sigue:

$$k_n = k_{n-1} - \frac{1}{\left(n - \frac{n-a}{2}\right)^2}$$

donde: $n > 3$,
 $k_3 = 1$, y
 $a = 2$, cuando n es par; ó $a = 3$ cuando n es impar.

De esta forma, el valor de los coeficientes determinados por fórmula se recogen en la **Tabla 2.22**:

Nº de revistas (n)	Valor de k_n según la fórmula
3	1,000
4	0,889
5	0,826
6	0,764
7	0,724
8	0,684
9	0,656
10	0,628
11	0,608
12	0,588

Tabla 2.22

Comparados los anteriores valores con los obtenidos empíricamente, Kuhn llegaba a la conclusión de que se podía generalizar la fórmula para determinar el coeficiente k_n . Pero no llegó a validar la fórmula cuando n era mayor que 10.

Habiendo seleccionado 180 combinaciones de soportes, comprobó que en ningún caso el error era mayor que el 1,8%, y que en 130 ocasiones el error era menor que el 1%. No obstante, hay que tener en cuenta que definía así el error:

$$\text{error (\%)} = \frac{\text{alcance neto estimado} - \text{alcance neto observado}}{\text{población}}$$

Este porcentaje de error es distinto del de Bower⁶⁵, como se explicaba antes. El definido Kuhn es menos exigente, de ahí que sea necesario comparar los modelos con los mismos tests, como vamos a hacer en nuestra aplicación.

Aunque las combinaciones tuvieran revistas de distintas categorías (de radio-televisión, de cine, de entretenimiento, etc) los resultados no variaban sustancialmente. Finalmente supone que la fórmula de Agostini y la suya propia, deberían aplicarse, sin necesidad de ser modificadas, a segmentos de la muestra.

2.2. Ejemplo

Nuevamente nos referimos a la selección de las cuatro revistas de la **Tabla 2.23** para el público objetivo nº 1:

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.23

⁶⁵ BOWER, John. "Net Audiences of U.S. and Canadian Magazines: Seven Tests of Agostini's Formula". *Journal of Advertising Research*. Marzo 1993. Vol. 3. Nº 1. Pág. 15.

Como $n = 4$, y sabiendo que k_{n-1} vale 1, y a toma el valor 2, podemos determinar el coeficiente k_n :

$$k_n = k_{n-1} - \frac{1}{\left(n - \frac{n-a}{2}\right)^2} = 1 - \frac{1}{\left(4 - \frac{4-2}{2}\right)^2} = 0,889$$

Determinando el valor de x y z :

$$x = \frac{D}{A} = \frac{55.208 + 47.445 + 30.036 + 23.935 + 29.574 + 11.555}{309.918 + 236.902 + 130.265 + 129.992} \approx 0,245$$

$$z = e^{-k_n \cdot x} = e^{-0,889 \cdot 0,245} \approx 0,804$$

podemos hallar finalmente el alcance neto C :

$$C = z \cdot A = 0,804 \cdot 807.077 \approx 649.046$$

2.3 Aplicación a nuestro estudio

El modelo de Kuhn ha producido en todos los casos sobrestimaciones en el alcance neto, un 13,6% de media con una dispersión del 8,2%. No se han podido determinar relación entre los distintos públicos objetivos y los errores (Tabla 2.24); aunque hay variaciones entre los distintos target, parecen tratarse de variaciones aleatorias. En cuanto al criterio de selección de soportes empleado, no se presentan siquiera variaciones entre las distintas alternativas.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
H, 14-34, A-MA	15,2376	9,0631	15,2376	9,0631
H, 14-34, MM	14,9528	8,8196	14,9528	8,8196
H, 14-34, MB-B	14,3797	8,3640	14,3797	8,3640
H, 35-54, A-MA	12,5803	7,2342	12,5803	7,2342
H, 35-54, MM	12,5908	7,0875	12,5908	7,0875
H, 35-54, MB-B	9,8220	4,9613	9,8220	4,9613
H, >54, A-MA	11,1379	6,9407	11,1379	6,9407
H, >54, MM	8,0633	4,8454	8,0633	4,8454
H, >54, MB-B	8,3938	4,7045	8,3938	4,7045
M, 14-34, A-MA	15,4311	9,2266	15,4311	9,2266
M, 14-34, MM	15,5259	9,1522	15,5259	9,1522
M, 14-34, MB-B	15,4741	8,7711	15,4741	8,7711
M, 35-54, A-MA	13,2479	8,0314	13,2479	8,0314
M, 35-54, MM	14,9895	8,9982	14,9895	8,9982
M, 35-54, MB-B	11,9365	6,6265	11,9365	6,6265
M, >54, A-MA	16,4741	10,3213	16,4741	10,3213
M, >54, MM	14,4086	8,3535	14,4086	8,3535
M, >54, MB-B	11,2618	6,3564	11,2618	6,3564
TOTAL	17,9888	8,6051	17,9888	8,6051
TODOS LOS CASOS	13,6161	8,2205	13,6161	8,2205

Tabla 2.24: Errores del modelo de Kuhn, según el público objetivo.

Igual que en los modelos anteriores, comprobamos que los errores se hacen mayores con el número de soportes de la campaña. Nótese que el aumento es mucho mayor que en el modelo de Agostini o en el posterior de Hofmans. Aquí las sobrestimaciones de las campañas con 20 soportes, son del 22,5%, como media (Tabla 2.25). Sin embargo, las sobrestimaciones con 4 revistas apenas alcanzan un 1,8%, debido a que este modelo se concibió a partir de un número limitado de revistas.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
4	1,8079	0,6928	1,8079	0,6928
8	8,1579	1,6450	8,1579	1,6450
12	14,1217	2,7895	14,1217	2,7895
16	18,9006	3,6328	18,9006	3,6328
20	22,5109	4,5061	22,5109	4,5061
TODOS LOS CASOS	13,6161	8,2205	13,6161	8,2205

Tabla 2.25: Errores del modelo de Kuhn, según el número de soportes.

3. La fórmula de Kaatz

3.1. Planteamiento original y adaptaciones

Kaatz⁶⁶ era consciente de que las investigaciones posteriores a Agostini habían encontrado en esa fórmula una persistente tendencia a la sobrestimación del alcance neto; también se veía que la sobrestimación aumentaba cuando era mayor el número de soportes de la combinación. Todo ello indicaba la necesidad de un factor de corrección que compensara esas desviaciones.

La investigación de Kaatz se basaba en las combinaciones de tres, cuatro y cinco publicaciones, recogidas en 1960 por Nielsen Media Service. Las publicaciones son doce revistas de consumo y suplementos de periódicos. Se trataba de 100 combinaciones elegidas al azar.

La media de error obtenido fue del 1,8%, y el coeficiente de correlación entre las estimaciones de la fórmula de Agostini y los datos obtenidos por NMS fue de 0,982. En el 8% de los casos hubo subestimaciones y en el 89% sobrestimaciones. El error de estimación se incrementaba con el número de publicaciones, desde un 1% con tríos hasta un 2,1% con quintetos. Los resultados son bastante paralelos a los obtenidos por Bower, Caffyn y Sagovsky, Marc y Kuhn: la fórmula de Agostini es un instrumento válido, pero necesita un factor de corrección.

El factor de corrección de Kaatz se deriva de la regresión lineal entre las variables Y y X, definidas como:

$$Y = \frac{\text{alcance neto observado}}{\text{alcance neto estimado por Agostini}}$$

$$X = \text{factor } x \text{ del modelo de Agostini}$$

⁶⁶ KAATZ, Ronald B. "Improving Agostini's Formula for Net Audience". *Journal of Advertising Research*. Septiembre 1963. Vol 3. Nº 3. Págs. 43-4.

Determinados los coeficientes en la regresión lineal, la ecuación quedó determinada de la siguiente manera:

$$Y_e = 1,006 - 0,053 \cdot X$$

A partir de esta ecuación, se obtiene el factor de ajuste del alcance neto en relación con el factor x de la fórmula de Agostini. Así se muestra en la **Tabla 2.26**:

Factor x de la fórmula de Agostini	Factor ajustado de la audiencia neta
0,150	0,998
0,200	0,995
0,250	0,993
0,300	0,990
0,350	0,987
0,400	0,985
0,450	0,982
0,500	0,980
0,550	0,977
0,600	0,974
0,650	0,972
0,700	0,969

Tabla 2.26

El factor ajustado, según el valor que tome x en cada caso, se multiplicará por el alcance neto estimado por Agostini, dando lugar a un alcance neto ajustado. Con este ajuste, Kaatz consiguió reducir la media de errores de las cien combinaciones estudiadas hasta el 1,1%. En el caso de los tríos el error era del 0,5% y con los quintetos del 1,2%. Ahora el error no tiene una dirección predominante, positiva o negativa, sino que aproximadamente las sobrestimaciones se compensan con las subestimaciones. No obstante se observa que el factor de ajuste no elimina el creciente error cuando es mayor el número de soportes.

Procedamos del mismo modo que Kaatz en nuestra investigación empírica. Primero debemos establecer la regresión entre la variable dependiente Y y la

independiente X. Para determinar Y se ha tomado como alcance observado, el del TOM2, y como alcance estimado, el de Agostini con $k = 1'048$. La regresión, realizada por el programa SPSS/PC+, se muestra en la salida de ordenador de la **Figura 2.6**:

REGRESSION /VARIABLES Y X /DESCRIPTIVES MEAN STDDEV CORR /DEPENDENT Y /METHOD ENTER.					
	Mean	Std Dev			
Y	1.003	0.021			
X	0.624	0.266			
N of Cases = 402					
Correlation:					
	Y	X			
Y	1.000	0.025			
X	0.025	1.000			
* * * * M U L T I P L E R E G R E S S I O N * * * *					
Multiple R		0.02456			
R Square		0.00060			
Adjusted R Square		-0.00190			
Standard Error		0.02089			
Analysis of Variance					
	DF	Sum of Squares		Mean Square	
Regression	1	0.00011		0.00011	
Residual	400	0.17457		0.00044	
F =	0.24150	Signif F =	0.6234		
----- Variables in the Equation -----					
Variable	B	SE B	Beta	T	Sig T
B	0.001925	0.003917	0.024564	0.491	0.6234
(Constant)	1.001610	0.002659		376.687	0.0000

Figura 2.6

Podemos observar a través de la media y de la desviación estándar que la variación de X es mayor que la de Y; dicho de otra forma, el error estimado es más constante que X (el cociente entre la suma de duplicaciones y el alcance bruto). También se observa en la **Figura 2.6** que la correlación entre X e Y es del 2,5%, prácticamente inexistente y contraria a los resultados de Kaatz. El modelo no explicaría ni en uno por ciento las desviaciones de la variable dependiente. Además el coeficiente de la variable dependiente no es significativo y el modelo en su conjunto tampoco lo es. Debemos, por tanto, rechazar este modelo pues se basa en premisas sin ninguna significación estadística.

Hemos aplicado el modelo de Kaatz, con dos modificaciones: por una parte, los parámetros resultantes de la regresión anterior; y por otra, la función del factor de ajuste no se considera en intervalos sino caso a caso (a cada valor de x le corresponde un valor del factor ajustado, en vez de que a un intervalo de x le corresponda un valor del factor ajustado). Esta segunda modificación no supone complejidad añadida con las hojas de cálculo, y parece más razonable, al tiempo que mejora ligeramente las estimaciones.

3.2. Ejemplo

Sin repetir los datos del ejemplo habitual, empezamos hallando el valor de x :

$$x = \frac{D}{A} = \frac{55.208 + 47.445 + 30.036 + 23.935 + 29.574 + 11.555}{309.918 + 236.902 + 130.265 + 129.992} \approx 0,245$$

De la regresión de la **Figura 2.6** habríamos obtenido la siguiente ecuación:

$$Y = 1,00161 - 0,001925 \cdot X$$

Sustituyendo nuestro valor de X , obtenemos el valor del factor de ajuste del alcance neto:

$$Y = 1,00161 - 0,001925 \cdot X = 1,00161 - 0,001925 \cdot 0,245 \approx 1,00114$$

Como $Y > 1$, este factor de ajuste va a corregir la subestimación del alcance neto de Agostini. Multiplicamos el alcance estimado por Agostini (con $k = 1'048$) por el factor de ajuste, estimando el alcance neto según Kaatz (C_K):

$$C_K = C_A \cdot Y = 642.102 \cdot 1,00114 \approx 642.833$$

3.3. *Test de la fórmula modificada*

El modelo de Kaatz, con la modificación realizada, se ha mostrado buen estimador del alcance neto, sin llegar a ser tan bueno como el de Agostini o el posterior de Hofmans. Presenta una media absoluta de errores del 1,6%, repartiéndose casi igualmente entre subestimaciones y sobrestimaciones. Aunque los errores varían entre los segmentos de población (**Tabla 2.27**), parece que lo hace aleatoriamente, sin tener relación con el sexo, la edad o la clase social. Como en el modelo de Hofmans que después veremos, nos encontramos con un *target* realmente atípico: los hombres, mayores de 54 años, de clase social media media; tiene un error absoluto dos veces y media superior al habitual; una vez comprobado que la recogida de datos y el cálculo han sido correctamente realizados, sólo nos queda afirmar que se trata de una diferencia atípica.

Considerado en sí mismo, el modelo de Kaatz ha fracasado en sus objetivos: recordemos que se trata de un ajuste de la fórmula de Agostini, para eliminar sus errores de estimación. El resultado ha sido que los errores se han agrandado ligeramente. Esto se ha debido a la falta de significación estadística de la regresión planteada por el autor, recogida en la **Figura 2.6**.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	1,2216	0,7892	1,3106	0,6210
H, 14-34, MM	1,0575	1,0663	1,2494	0,8192
H, 14-34, MB-B	0,7357	1,0282	1,0805	0,6308
H, 35-54, A-MA	-0,7376	1,1991	1,1152	0,8375
H, 35-54, MM	-0,4731	1,2589	1,1283	0,6929
H, 35-54, MB-B	-1,9382	2,2306	2,2572	1,8886
H, >54, A-MA	-1,5478	1,1391	1,6276	1,0154
H, >54, MM	-4,1586	2,5090	4,1805	2,4704
H, >54, MB-B	-2,0941	1,4965	2,2156	1,2993
M, 14-34, A-MA	1,4824	0,7161	1,4824	0,7161
M, 14-34, MM	1,6105	0,7603	1,6265	0,7236
M, 14-34, MB-B	1,6146	0,6225	1,6146	0,6225
M, 35-54, A-MA	-0,2624	1,2081	1,1525	0,3659
M, 35-54, MM	1,0990	0,5879	1,1621	0,4421
M, 35-54, MB-B	-1,2866	1,6972	1,8401	1,0291
M, >54, A-MA	2,4416	1,1919	2,4416	1,1919
M, >54, MM	0,8322	0,7518	0,9219	0,6322
M, >54, MB-B	-1,2266	1,2586	1,6449	0,5653
TOTAL	-1,1128	1,1818	1,4960	0,6102
TODOS LOS CASOS	-0,1973	2,0442	1,6514	1,2181

Tabla 2.27: Errores del modelo de Kaatz, según el público objetivo.

En este modelo volvemos a comprobar que ha sido indiferente el criterio de selección de soportes. No influye en la calidad de la estimación, como se refleja en la **Tabla 2.28**. Los criterios de la rentabilidad, la cobertura, la combinación de ambos o la afinidad, producen errores con la misma dimensión, signo y dispersión.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-0,0344	1,9679	1,5907	1,1485
cobertura	-0,3166	2,1751	1,8183	1,2222
rentabilidad+cobertura	-0,1214	2,1089	1,7367	1,1903
afinidad	-0,3356	1,9104	1,4301	1,3021
TODOS LOS CASOS	-0,1973	2,0442	1,6514	1,2181

Tabla 2.28: Errores del modelo de Kaatz, según el criterio de selección de soportes.

Más interesante resulta el análisis de la relación entre los errores y el número de soportes (**Tabla 2.29**). Para 4 y 8 revistas, el modelo de Kaatz da sobrestimaciones, pero

con más revistas pasan a ser subestimaciones. El error se incrementa a medida que aumentan los vehículos considerados, tanto en su valor medio como en su dispersión; como en otros modelos, la estimación es más precisa con pocos soportes. La diferencia del error medio absoluto es 1,1% entre la primera y la última alternativa (2,2%-1,1%). Esta evolución de los errores es semejante a la de Agostini, con el 1,2% (Tabla 2.7). Sin embargo, es menor que el 1,9% obtenido por Hofmans, cuyas estimaciones serán muy precisas con pocos soportes, pero con muchos soportes más imprecisas que los demás modelos. Los modelos de Agostini y Kaatz presentan menos variación que el de Hofmans cuando aumentan el número de soportes.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	0,9923	0,7953	1,0973	0,6405
8	0,2457	1,4479	1,2026	0,8319
12	-0,2393	2,0660	1,7576	1,0934
16	-0,5808	2,3355	1,9651	1,3721
20	-0,9659	2,6251	2,2424	1,6554
TODOS LOS CASOS	-0,1973	2,0442	1,6514	1,2181

Tabla 2.29: Errores del modelo de Kaatz, según el número de soportes.

4. Modelo de Hofmans

4.1. Fundamentos e hipótesis

Hofmans propuso un modelo⁶⁷ para determinar el número total de lectores, espectadores y oyentes que están expuestos a una combinación de medios. Puede ser usado para toda la población o un segmento concreto. Para llegar a su fórmula, plantea el caso de tener sólo dos soportes, con audiencias A_1 y A_2 , su duplicación D_{12} y audiencia neta C_2 . El alcance neto se define como:

$$C_2 = A_1 + A_2 - D_{12}$$

Según la fórmula de Agostini valdría:

$$C_2 = \frac{A_1 + A_2}{1 + k \cdot \frac{D_{12}}{A_1 + A_2}}$$

operando,

$$k = \frac{(A_1 + A_2 - C_2) \cdot (A_1 + A_2)}{D_{12} \cdot C_2}$$

y como la duplicación se puede definir como:

$$D_{12} = A_1 + A_2 - C_2$$

k tomará el siguiente valor:

$$k = \frac{A_1 + A_2}{C_2}$$

Por tanto, el parámetro k de Agostini no es constante, sino que varía según de la pareja de soportes. Para dos vehículos cualesquiera i y j , se generaliza así:

⁶⁷ HOFMANS, Pierre. "Measuring the Cumulative Net Coverage of Any Combination of Media". *Journal of Marketing Research*. Agosto de 1966. Vol. 3. Págs. 269-78.

$$k_{ij} = \frac{A_i + A_j}{C_{ij}}$$

Ahora no se trata multiplicar la suma de las duplicaciones por una k constante (como la fórmula de Agostini) sino de hallar la suma de las diferentes duplicaciones multiplicadas por su correspondiente k_{ij} . Analíticamente la nueva fórmula se expresa así:

$$C = \frac{\left(\sum_{i=1}^n A_i \right)^2}{\sum_{i=1}^n A_i + \sum_{i=1, i < j}^n k_{ij} D_{ij}}$$

La hipótesis que reemplaza la constante k por una variable dependiente de los pares de soportes, fue verificada empleando revistas norteamericanas, francesas y belgas. En seis combinaciones la fórmula de Hofmans nunca superaba un error del 1% mientras que la de Agostini llegaba hasta un 4%; además, en los seis casos, la estimación de Hofmans era más correcta.

A partir de los estudios de Politz⁶⁸, Hofmans plantea una nueva hipótesis: los valores de k aumentan cuando la combinación consta de más soportes. Para ello muestra la media de los valores de k según el número de revistas (véase la **Tabla 2.30**), aunque se han considerado pocos casos:

<i>Media de k</i>	<i>Nº de revistas</i>	<i>Nº de combinaciones</i>
1,251	2	10
1,254	3	10
1,262	4	5
1,271	5	1

Tabla 2.30

⁶⁸ ALFRED POLITZ RESEARCH, INC. "A Study of Four Media- A study of Duplication". New York. Time, Inc, 1953.

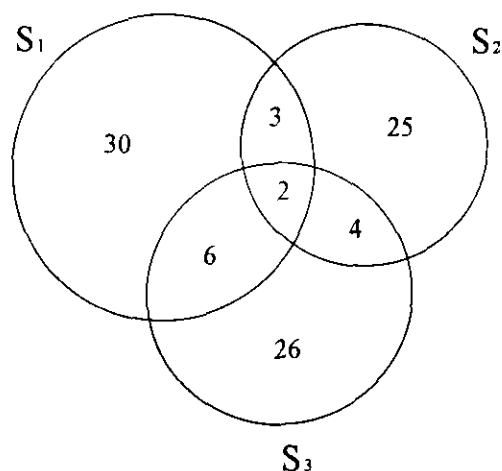
Como señalan Claycamp y McClelland⁶⁹, la mejora de Hofmans puede reducir los errores que se deriven de la no constancia del parámetro k en la fórmula de Agostini, aunque no suponga una garantía en todos los casos. Indican que está sustituyendo la constante k por una media ponderada k' , como sigue:

$$k' = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij} \cdot D_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n D_{ij}}$$

Por álgebra sabemos que el alcance neto es la suma de las audiencias individuales (A), menos la suma de las duplicaciones (D), más la suma de las triplicaciones (T), menos la suma de las cuadruplicaciones (Q),..., más/menos la suma de las duplicaciones de orden n (N).

$$C = A - D + T - Q + \dots \pm N$$

Para ilustrar dicha fórmula veamos un ejemplo con tres soportes, S_1 , S_2 y S_3 , representados gráficamente en la **Figura 2.7**:



Las audiencias individuales A_i correspondientes a los S_i soportes son:

$$A_1 = 30 + 3 + 2 + 6 = 41$$

$$A_2 = 25 + 3 + 2 + 4 = 34$$

$$A_3 = 26 + 6 + 2 + 4 = 38$$

Luego:

$$A = A_1 + A_2 + A_3 = 41 + 34 + 38 = 113$$

Figura 2.7

Las duplicaciones D_{ij} entre el

⁶⁹CLAYCAMP, H.J. and McCLELLAND, C.W. "Estimating Reach and the Magic of K ". *Journal of Advertising Research*. Junio 1968. Vol. 8. Nº 2. Pág. 48.

soporte S_i y el S_j valen:

$$D_{12} = 3 + 2 = 5$$

$$D_{13} = 6 + 2 = 8$$

$$D_{23} = 2 + 4 = 6$$

Luego: $D = D_{12} + D_{13} + D_{23} = 5 + 8 + 6 = 19$. Y la triplicación T_{123} es igual a 2. La audiencia neta, los que están al menos en uno de esos conjuntos, se puede hallar como sigue:

$$\begin{aligned} A_1 \cup A_2 \cup A_3 &= A_1 + A_2 + A_3 - (A_1 \cap A_2 + A_1 \cap A_3 + A_2 \cap A_3) + A_1 \cap A_2 \cap A_3 = \\ &= A_1 + A_2 + A_3 - (D_{12} + D_{13} + D_{23}) + T_{123} = \\ &= A - D + T = 113 - 19 + 2 = 96 \end{aligned}$$

Chandon⁷⁰ señala que la fórmula de Agostini, lo mismo que la de Hofmans, es una serie geométrica infinita de base A y con razón $k \cdot D/A$:

$$C = \frac{A}{1 + k \cdot \frac{D}{A}} = A - k \cdot D + \frac{(k \cdot D)^2}{A} - \frac{(k \cdot D)^3}{A^2} + \dots \pm \frac{(k \cdot D)^n}{A^{n-1}} \pm \dots$$

Se pueden comparar término a término las ecuaciones anteriores, de manera que el primer término de la serie sea igual a la suma de audiencias individuales, el segundo igual a la suma de duplicaciones, el tercero a la suma de triplicaciones, etc:

$$D = k \cdot D ; \quad T = \frac{(k \cdot D)^2}{A} ; \quad Q = \frac{(k \cdot D)^3}{A^2} ; \quad \dots ; \quad N = \frac{(k \cdot D)^n}{A^{n-1}}$$

Chandon define también el rango que tiene el parámetro k_{ij} de la fórmula de Hofmans. Cuando no exista duplicación entre dos soportes, k_{ij} será igual a 1. Al contrario, cuando todos los lectores lean ambos soportes, $D_{ij} = A_i = A_j$, entonces $k_{ij} = 2$.

⁷⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 318-20.

Por tanto, el parámetro k_{ij} variará entre 1 y 2. La audiencia neta, estimada por Hofmans, estará limitada por los casos extremos de no existencia de duplicación o de duplicación perfecta:

$$A - D + \frac{D^2}{A} - \frac{D^3}{A^2} + \dots \pm \frac{D^n}{A^{n-1}} \pm \dots \leq C \leq A - 2D + \frac{(2D)^2}{A} - \frac{(2D)^3}{A^2} + \dots \pm \frac{(2D)^n}{A^{n-1}} \pm \dots$$

Del mismo modo, la suma de las duplicaciones de orden r , que notaremos como R_r , estará limitada superior e inferiormente:

$$\frac{D^{r-1}}{A^{r-2}} \leq R_r \leq \frac{(2D)^{r-1}}{A^{r-2}}$$

En España, Barreiro⁷¹ realizó un programa informático para la aplicación de la fórmula de Hofmans.

4.2. Ejemplo

Aplicaremos la fórmula de Hofmans a nuestro caso, cuyos datos volvemos a recoger en la **Tabla 2.31**:

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.31

Tendremos tantos valores de k_{ij} como duplicaciones entre las revistas:

⁷¹ BARREIRO FERNÁNDEZ, Jose Manuel. "Las duplicaciones de audiencias entre soportes publicitarios: desarrollo de un modelo informático de la fórmula de Hofmans para la determinación de coberturas y distribuciones de exposiciones de una combinación de soportes". Ponencia presentada en los III Encuentros de Profesores Universitarios de Marketing. Salamanca. Septiembre 1991.

$$k_{12} = \frac{A_1 + A_2}{C_{12}} = \frac{A_1 + A_2}{A_1 + A_2 - D_{12}} = \frac{309.918 + 236.902}{309.918 + 236.902 - 55.208} \approx 1,1123$$

$$k_{13} = \frac{A_1 + A_3}{C_{13}} = \frac{A_1 + A_3}{A_1 + A_3 - D_{13}} = \frac{309.918 + 130.265}{309.918 + 130.265 - 47.455} \approx 1,1208$$

$$k_{14} = \frac{A_1 + A_4}{C_{14}} = \frac{A_1 + A_4}{A_1 + A_4 - D_{14}} = \frac{309.918 + 129.992}{309.918 + 129.992 - 30.036} \approx 1,0733$$

$$k_{23} = \frac{A_2 + A_3}{C_{23}} = \frac{A_2 + A_3}{A_2 + A_3 - D_{23}} = \frac{236.902 + 130.265}{236.902 + 130.265 - 23.935} \approx 1,0697$$

$$k_{24} = \frac{A_2 + A_4}{C_{24}} = \frac{A_2 + A_4}{A_2 + A_4 - D_{24}} = \frac{236.902 + 129.992}{236.902 + 129.992 - 29.574} \approx 1,0877$$

$$k_{34} = \frac{A_3 + A_4}{C_{34}} = \frac{A_3 + A_4}{A_3 + A_4 - D_{34}} = \frac{130.265 + 129.992}{130.265 + 129.992 - 11.555} \approx 1,0465$$

Ya podemos determinar el alcance neto de las cuatro revistas:

$$\sum_{i=1}^n A_i = 309.918 + 236.902 + 130.265 + 129.992 = 807.077$$

$$\begin{aligned} \sum_{i=1, i < j}^n k_{ij} D_{ij} &= 1,1123 \cdot 55.208 + 1,1208 \cdot 47.455 + 1,0733 \cdot 30.036 + \\ &+ 1,0697 \cdot 23.935 + 1,0877 \cdot 29.574 + 1,0465 \cdot 11.555 = 216.696 \end{aligned}$$

$$C = \frac{\left(\sum_{i=1}^n A_i \right)^2}{\sum_{i=1}^n A_i + \sum_{i=1, i < j}^n k_{ij} D_{ij}} = \frac{(807.077)^2}{807.077 + 216.696} \approx 636.185 \quad [72]$$

⁷² La diferencia entre la solución exacta y este resultado se debe a que, en la presentación de estos ejemplos, no se indican los decimales. Los resultados (cfr. campaña nº 10 del fichero **RTDOS.WK1**) se han calculado con gran número de decimales.

4.3. Aplicación a la investigación propuesta

El modelo de Hofmans presenta una subestimación media pequeña, en torno al 1,4%. Obviando el signo de errores, podemos decir que la dimensión media de los errores se sitúa en 1,8%, con una desviación típica del 1,5%. Se trata de un margen pequeño que hace que este modelo se pueda considerar bastante preciso. No obstante resulta chocante que no mejore el modelo de Agostini, como se venía suponiendo en las anteriores investigaciones. Vemos que poco puede haber influido el criterio de selección de soportes (Tabla 2.32) pues los errores son bastante similares en todas las alternativas.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-1,1969	1,8292	1,6422	1,4386
cobertura	-1,7658	1,8606	2,0477	1,5415
rentabilidad+cobertura	-1,4518	1,8841	1,8087	1,5411
afinidad	-1,3418	2,0763	1,8965	1,5794
TODOS LOS CASOS	-1,4424	1,9137	1,8471	1,5257

Tabla 2.32: Errores del modelo de Hofmans, según el criterio de selección de soportes.

En cuanto a los públicos objetivo, se observan diferencias significativas entre las distintas alternativas, pero parece un fenómeno aleatorio (cfr. Tabla 2.33). Ni el sexo, ni la edad ni la clase social parecen explicar el comportamiento de los errores. Un segmento tiene diferencias particularmente significativas, con unas subestimaciones del 4,8%: los hombres, mayores de 54 años y de clase social media media. Sólo encontramos dos posibles explicaciones: (1) fallos en la recogida de información o en las operaciones aritméticas, o (2) estamos ante un valor extremo de la distribución. Lo primero queda descartado tras revisar concienzudamente la recogida de datos, su tratamiento informático y el proceso de cálculo. Por ello, deducimos que no tiene otra explicación más que la aleatoriedad.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	0,0170	0,9120	0,7099	0,5491
H, 14-34, MM	-0,1428	1,1437	0,8997	0,6910
H, 14-34, MB-B	-0,2877	0,8547	0,6648	0,5936
H, 35-54, A-MA	-1,6130	1,3115	1,7590	1,0963
H, 35-54, MM	-1,3675	1,1900	1,4346	1,1036
H, 35-54, MB-B	-2,4384	2,1098	2,5145	2,0135
H, >54, A-MA	-2,4387	1,3075	2,5164	1,1421
H, >54, MM	-4,8566	2,4732	4,8566	2,4732
H, >54, MB-B	-2,3249	1,4400	2,3426	1,4096
M, 14-34, A-MA	1,4289	2,1491	2,1687	1,3479
M, 14-34, MM	-0,3853	0,6889	0,6557	0,4224
M, 14-34, MB-B	-0,6976	0,5099	0,6995	0,5071
M, 35-54, A-MA	-2,5996	1,4021	2,6416	1,3167
M, 35-54, MM	-0,7757	0,7406	0,8213	0,6868
M, 35-54, MB-B	-2,5239	1,6362	2,5488	1,5950
M, >54, A-MA	0,4415	1,4557	1,3630	0,6074
M, >54, MM	-1,3987	0,6287	1,3987	0,6287
M, >54, MB-B	-2,3908	1,2279	2,4676	1,0560
TOTAL	-2,2093	1,0091	2,2207	0,9831
TODOS LOS CASOS	-1,4424	1,9137	1,8471	1,5257

Tabla 2.33: Errores del modelo de Hofmans, según el público objetivo.

Como en los modelos anteriores, comprobamos que los errores aumentan cuando la combinación incluye más soportes. En valor absoluto, la media de errores ha ido creciendo progresivamente, desde el 0,6% de 4 revistas hasta cuadruplicarse con 20 revistas (**Tabla 2.34**). Crece igualmente la desviación típica, es decir, se hace mayor la dispersión respecto a la media. Si comparamos estos resultados con los equivalentes del método de Agostini (**Tabla 2.7**), encontramos que Hofmans estima mejor con pocos soportes y Agostini mejor con más revistas; en conjunto, Agostini produce menos errores.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
4	-0,2673	0,8857	0,6390	0,6656
8	-1,1954	1,5348	1,5740	1,1370
12	-1,6177	1,9334	1,9972	1,5322
16	-1,7531	2,1782	2,2771	1,6138
20	-2,0299	2,3868	2,5451	1,8188
TODOS LOS CASOS	-1,4424	1,9137	1,8471	1,5257

Tabla 2.34: Errores del modelo de Hofmans, según el número de soportes.

5. Modelo de Starch

5.1. Hipótesis y desarrollo

Starch⁷³ propuso un método que va sumando el alcance neto aportado por los diversos soportes de la campaña. Al soporte de mayor audiencia se le agregará el alcance neto añadido por los demás soportes. Como señala Durand⁷⁴, el método de Starch se basa en los modelos probabilísticos, estableciendo hipótesis para los soportes de este tipo:

$$P(S_1 \cap S_2 \cap S_3) = P(S_1 \cap S_2) \cdot \frac{P(S_2 \cap S_3) + P(S_1 \cap S_3)}{P(S_1) + P(S_2)}$$

Está planteando, por tanto, unas hipótesis que permiten estimar las triplicaciones (o en su caso, las repeticiones de mayor orden) en función de las duplicaciones, simplificándose el cálculo del alcance neto de n soportes.

Starch admitía la falta de precisión que se produce cuando se considera un número elevado de soportes, recomendando en esos casos que se realicen tabulaciones especiales a partir de las fichas individuales de encuestas. El modelo se aplica en las siguientes etapas:

1. Colocar las revistas en orden de preferencia.

2. Suponiendo que al primer soporte se le van agregando los otros soportes, sumar las audiencias de los soportes agregados a las anteriores y restarles la duplicación correspondiente: $[A_2 + A_1 - D_{21}]$ $[A_3 + A_1 - D_{31}]$ $[A_3 + A_2 - D_{32}]$ $[A_4 + A_1 - D_{41}]$ $[A_4 + A_2 - D_{42}]$ $[A_4 + A_3 - D_{43}]$.

⁷³ STARCH, Daniel and STAFF, Inc. *Starch Magazine Report*. New York, 1967.

⁷⁴ DURAND, Jacques. "L'analyse de la duplication entre supports de publicité et ses enseignements pour la théorie statistique". *Gestion*. Diciembre 1965.

3. Se determina el porcentaje de los no duplicados en cada soporte sobre su audiencia: $[(A_2 - D_{21})/A_2]$ $[(A_3 - D_{31})/A_3]$ $[(A_3 - D_{32})/A_3]$ $[(A_4 - D_{41})/A_4]$ $[(A_4 - D_{42})/A_4]$ $[(A_4 - D_{43})/A_4]$.

4. Se acumulan los anteriores porcentajes, multiplicando las combinaciones resultantes al incorporar un nuevo soporte:

$$\begin{aligned} & (A_2 - D_{21}) / A_2 \\ & [(A_3 - D_{31}) / A_3] \cdot [(A_3 - D_{32}) / A_3] \\ & [(A_4 - D_{41}) / A_4] \cdot [(A_4 - D_{42}) / A_4] \cdot [(A_4 - D_{43}) / A_4] \end{aligned}$$

5. Se determina el alcance neto añadido por el nuevo soporte, multiplicando su audiencia total por el ratio anterior correspondiente.

$$A.N.A.(A_2) = A_2 \cdot \{(A_2 - D_{21}) / A_2\}$$

$$A.N.A.(A_3) = A_3 \cdot [(A_3 - D_{31}) / A_3] \cdot [(A_3 - D_{32}) / A_3]$$

$$A.N.A.(A_4) = A_4 \cdot [(A_4 - D_{41}) / A_4] \cdot [(A_4 - D_{42}) / A_4] \cdot [(A_4 - D_{43}) / A_4]$$

6. El alcance neto total se halla sumando la parte de audiencia no duplicada de los nuevos soportes, es decir, sumando a la audiencia del primer soporte los alcances netos añadidos que van aportando los soportes entrantes:

$$C_n(A_1, A_2, A_3, A_4) = A_1 + A.N.A.(A_2) + A.N.A.(A_3) + A.N.A.(A_4)$$

5.2. Ejemplo

Volvemos a utilizar la combinación de cuatro soportes a la que nos hemos referido anteriormente: en la **Tabla 2.35** están recogidas las audiencias individuales y las duplicaciones de las cuatro revistas.

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.35

Tomamos la revista *Muy Interesante* como primer soporte. Los alcances netos añadidos por los soportes entrantes son los siguientes:

- $A.N.A.(Teleprograma) = A_2 \cdot (A_2 - D_{21}) / A_2 = 236.902 \cdot (236.902 - 55.208) / 236.902 = 181.694$
- $A.N.A.(El Jueves) = A_3 \cdot [(A_3 - D_{31}) / A_3] \cdot [(A_3 - D_{32}) / A_3] = 130.265 \cdot [(130.265 - 47.445) / 130.265] \cdot [(130.265 - 23.935) / 130.265] = 67.603$
- $A.N.A.(Teleindiscreta) = A_4 \cdot [(A_4 - D_{41}) / A_4] \cdot [(A_4 - D_{42}) / A_4] \cdot [(A_4 - D_{43}) / A_4] = 129.992 \cdot [(129.992 - 30.036) / 129.992] \cdot [(129.992 - 29.574) / 129.992] \cdot [(129.992 - 11.555) / 129.992] = 70.352$

Finalmente, el alcance neto total será la suma de la audiencia de *Muy Interesante* más los alcances añadidos anteriores:

$$C = A_1 + A.N.A.(Teleprograma) + A.N.A.(El Jueves) + A.N.A.(Teleindiscreta) = 309.918 + 181.694 + 67.603 + 70.352 = 629.567$$

5.3. Tests sobre las campañas propuestas

Al aplicar el modelo de Starch a las 402 campañas, hemos encontrado que subestima el alcance neto un 5,5% por término medio, oscilando principalmente los errores entre un 2,5% y un 9,3%. Los errores son negativos en casi todos los casos (subestimación) y por eso, al ser considerados en valor absoluto, la magnitud no varía

apenas. Dada la desviación general de todos los casos, se podría plantear un factor que corrigiera las subestimaciones sistemáticas. Pero antes hemos de analizar los posibles determinantes de estos errores. Primero hemos buscado la relación con el segmento de población (véase **Tabla 2.36**), pareciendo que los errores varían aleatoriamente entre los distintos *target*.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
H, 14-34, A-MA	-5,0056	3,2799	5,0056	3,2799
H, 14-34, MM	-4,1355	2,4260	4,1355	2,4260
H, 14-34, MB-B	-4,8121	2,5974	4,8121	2,5974
H, 35-54, A-MA	-6,3621	4,2705	6,3679	4,2614
H, 35-54, MM	-4,0730	2,8147	4,0871	2,7931
H, 35-54, MB-B	-3,9852	2,4116	3,9852	2,4116
H, >54, A-MA	-5,4265	2,9203	5,4768	2,8198
H, >54, MM	-7,5874	3,7690	7,5874	3,7690
H, >54, MB-B	-3,5022	2,0201	3,5022	2,0201
M, 14-34, A-MA	-6,1774	4,2924	6,1774	4,2924
M, 14-34, MM	-6,3863	3,5453	6,3863	3,5453
M, 14-34, MB-B	-6,3317	3,4129	6,3317	3,4129
M, 35-54, A-MA	-8,5942	4,8697	8,6233	4,8153
M, 35-54, MM	-6,7163	3,2366	6,7163	3,2366
M, 35-54, MB-B	-5,8596	3,1165	5,8596	3,1165
M, >54, A-MA	-5,8495	1,9725	5,8495	1,9725
M, >54, MM	-7,1401	2,7142	7,1401	2,7142
M, >54, MB-B	-5,1512	2,3346	5,1512	2,3346
TOTAL	-7,6363	3,4024	7,6363	3,4024
TODOS LOS CASOS	-5,9270	3,4444	5,9319	3,4359

Tabla 2.36: Errores del modelo de Starch, según el público objetivo.

Parece que tampoco existe relación entre el criterio de selección de soportes y los errores, ni en cuanto a su magnitud, signo ni dispersión (cfr. **Tabla 2.37**).

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
rentabilidad	-5,5943	3,3687	5,6051	3,3506
cobertura	-6,5227	3,4855	6,5227	3,4855
rentabilidad+cobertura	-6,0476	3,4235	6,0476	3,4235
afinidad	-5,4836	3,4550	5,4932	3,4396
TODOS LOS CASOS	-5,9270	3,4444	5,9319	3,4359

Tabla 2.37: Errores del modelo de Starch, según el criterio de selección de soportes.

En cuanto al número de soportes de la campaña, sí encontramos una clara relación con los errores: con 4 revistas la subestimación es del 1,1% y va creciendo progresivamente hasta ser de un 9,1% con 20 revistas (**Tabla 2.38**). Como en el modelo anterior, no recogemos todas las alternativas para que resulte más fácil la lectura de los datos. Se puede deducir que incrementarse el número de soportes, el modelo de Starch produce subestimaciones mayores y es menos preciso, como indica la dispersión de los errores medida por la desviación estándar. Más que buscar un factor que corrija las subestimaciones sistemáticas de este modelo, pensamos que lo adecuado sería emplearlo únicamente cuando el número de soportes es pequeño (nótese que, para 4 soportes, la media de errores absolutos es muy parecida a la del modelo de Agostini).

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
4	-1,1075	0,9196	1,1340	0,8863
8	-4,0702	1,6714	4,0702	1,6714
12	-6,3345	1,9345	6,3345	1,9345
16	-7,9552	2,0189	7,9552	2,0189
20	-9,1856	2,1657	9,1856	2,1657
TODOS LOS CASOS	-5,9270	3,4444	5,9319	3,4359

Tabla 2.38: Errores del modelo de Starch, según el número de soportes.

6. Estimación de duplicaciones por Little y Lodish

6.1. Fundamentos e hipótesis

Little y Lodish⁷⁵ hicieron un modelo de planificación integral (véase **IV Parte, §B.6**) que entiende la duplicación a partir de una aproximación estocástica, corregida por una constante determinada empíricamente. Admiten la hipótesis de que la duplicación de audiencia es aleatoria, de forma que la probabilidad de estar expuesto a dos soportes es igual al producto de las probabilidades de exposición de cada soporte. Corrigen el error resultante, manteniendo que la probabilidad de exposición conjunta está en función del producto de probabilidades individuales:

$$P(S_i \cap S_j) = f(P(S_i) \cdot P(S_j))$$

Si se encontrara una relación funcional constante, no necesitaríamos acceder a las duplicaciones, con lo que se facilitaría mucho la búsqueda de datos; simplemente con las audiencias de cada soporte, podríamos hallar las duplicaciones y, usando otro de los modelos comentados en esta parte, determinar la audiencia neta. Los autores encontraron una sencilla ecuación para los datos de Simmons⁷⁶:

$$P(S_i \cap S_j) = [P(S_i) \cdot P(S_j)]^{0,65}$$

La relación es constante cualquiera que sea el soporte escogido. El parámetro $\alpha = 0,65$ no es más que la medida de la correlación positiva entre las audiencias de esos soportes. Los autores reconocen que su fórmula produce un error medio del 1,7%, y cuando se trata de la media de errores absolutos un 10,7%.

⁷⁵ LITTLE, John .D.C. and LODISH, Leonard M.: "A Media Planning Calculus". *Operation Research*. Enero-Febrero 1969. Págs. 18-20.

⁷⁶ SIMMONS, W.R. and ASSOCIATES RESEARCH, INC. *Mass Markets and the Media Reaching Them*. 1966 Magazine Reach and Frequency Report. New York. 1966.

Cuando Chandon⁷⁷ aplicó el modelo de Lodish y Little a 55 pares de revistas, para tres segmentos de población⁷⁸, concluyó que sus resultados eran muy pobres y que la constante α debería estimarse separadamente para cada *target*.

6.2. Test de las duplicaciones estimadas

Hemos buscado una ecuación que explique nuestras duplicaciones, a partir del modelo de Little y Lodish. Tenemos 8.265 duplicaciones, de 30 revistas referidas a 19 segmentos de población $[19 \cdot (30-29)/2]$. La regresión se ha realizado con el Excel 3.0, porque el SPSS/PC+ no admitía esa magnitud de datos.

Inicialmente la forma geométrica se ha transformado en una ecuación lineal tomando logaritmos. Además por comodidad no se expresa en términos probabilísticos, que sería lo mismo dividido por la población objetivo:

$$P(S_i \cap S_j) = [P(S_i) \cdot P(S_j)]^\alpha \Rightarrow D_{ij} = \left[\frac{A_i \cdot A_j}{P} \right]^\alpha \Rightarrow \ln(D_{ij}) = \alpha \cdot \ln \left[\frac{A_i \cdot A_j}{P} \right]$$

Dicha fórmula se trata de un modelo lineal, donde la variable independiente $\ln(D_{ij})$ queda explicada por la variable dependiente $\ln \left[\frac{A_i \cdot A_j}{P} \right]$ con su correspondiente coeficiente α , sin término independiente. En la **Tabla 2.39** se muestran los resultados de la regresión:

⁷⁷ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 338-41.

⁷⁸ ALFRED POLITZ MEDIA STUDIES. "Main Report 1964 Magazine Study: Audience of 11 Magazines". Febrero 1965. New York.

α	1,062594779	Error estándar α	0,002371763
r^2	0,768235511	Error estándar var. ind.	1,467277526
F	27.011,69281	Grados de libertad	8.149
Suma de los cuadrados de la regresión	58.153,56361	Suma de los cuadrados de los residuos	17.544,0093

Tabla 2.39

Por tanto el modelo resultante sigue la siguiente expresión:

$$D_{ij} = \left[\frac{A_i \cdot A_j}{P} \right]^{1,0626}$$

Es un modelo que explica el 77% de la variación de las duplicaciones, como indica que el coeficiente de determinación r^2 sea igual a 0,77. El modelo es significativo en su conjunto, dado el valor del estadístico F. Los grados de libertad, número de datos menos número de coeficientes, es de 8.149: en total se han empleado 8.150 datos; los 115 restantes, hasta las 8.265 duplicaciones, han sido rechazados porque bien $A_i=0$ (32 casos) ó $A_j=0$ (82 casos) ó ambas valen 0 (1 caso); es evidente que esos casos se pueden omitir en la regresión porque, si la audiencia de una revista es nula o lo son ambas, la duplicación será nula. Sin embargo, para hacer la regresión, la duplicación no podía ser nula, porque su logaritmo (en cualquier base) no estaría definido. Por esta misma razón cuando no existía duplicación (en 983 casos) se ha supuesto que hubiera una persona que estuviera expuesta a los dos soportes ($D_{ij}=1$): esta variación, que produce un error despreciable, es necesario para poder determinar el $\ln(D_{ij})$.

Aunque de los resultados se concluye que el modelo está bien fundamentado, no es suficiente para determinar con precisión cada duplicación. Hemos querido analizar algunos errores de estimación, tras definirse como:

$$error = \frac{\text{duplicación estimada} - \text{duplicación real}}{\text{duplicación real}} \cdot 100$$

Se seleccionan los errores en las 30 revistas, sobre el público objetivo total. Como se puede observar en la **Tabla 2.40** existe una gran diversidad en la media de errores encontrados entre cada revista, lo que refuta la constancia defendida por los autores del modelo. Algunos errores son lo suficientemente elevados como para desechar el intento de Little y Lodish: estimar la duplicación a partir de las audiencias individuales, supone un riesgo muy elevado. Obvio es decir, que esas duplicaciones estimadas no se podrían emplear entonces para el cálculo de la audiencia neta. No parece que se pueda establecer una relación entre las medias de errores y el contenido de las revistas (información general, TV, revistas del corazón,...). Quizá posteriores análisis puedan significar estadísticamente diferencias que parecen existir en los hábitos de lectura.

Nº de Revista	Revista	Media de errores	Nº de casos
1	Teleprograma	34,02%	29
2	Teleindiscreta	32,83%	29
3	Pronto	41,75%	29
4	Muy Interesante	9,00%	29
5	Hola	13,51%	29
6	Semana	11,26%	29
7	TV Plus	-1,20%	29
8	Lecturas	2,00%	29
9	El Jueves	-1,04%	29
10	Cruz Roja	13,00%	29
11	Diez Minutos	-5,43%	29
12	Nuevo Vale	-0,32%	29
13	Clan TV	4,02%	29
14	Interviú	2,67%	29
15	Burda	27,81%	29
16	Super Pop	-7,03%	29
17	Mía	-15,98%	29
18	Solomoto 30	-2,80%	29
19	Natura	-2,03%	29
20	Más Allá	-17,53%	29
21	Motociclismo	-3,80%	29
22	Autopista	1,19%	29
23	Ragazza	-22,82%	29
24	Fotogramas	-1,83%	29
25	Tiempo	-2,97%	29
26	Cambio 16	-3,21%	29
27	Motor 16	-14,94%	29
28	Labores Hogar	12,89%	29
29	Solo Moto Actual	-11,08%	29
30	Patrones	8,80%	29
	TODAS	3,36%	435

Tabla 2.40

6.3. Fórmulas alternativas

Se ha intentado mejorar el modelo de Little y Lodish, variando la estructura de la ecuación estimada. Primeramente hemos querido dibujar la nube de puntos en la **Figura 2.8**, y así poder escoger la estructura más adecuada. Hay que notar que el programa Excel 3.0 sólo nos ha permitido dibujar los primeros 4.000 puntos.

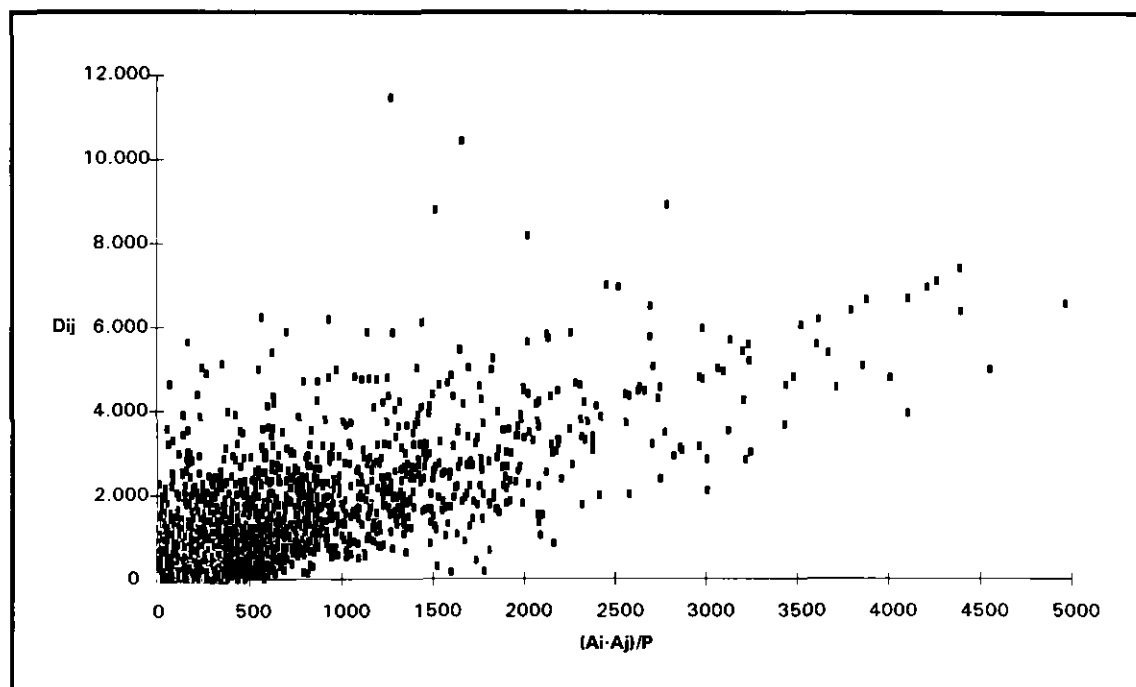


Figura 2.8

La nube de esos 4.000 puntos muestra más dispersión de la que realmente existe, pues figuran todos los puntos extremos, mientras que los más comunes se difuminan al superponerse. Con esta distribución de los datos hemos planteado varias alternativas de ajuste, que se recogen a continuación.

6.3.1. Ecuación geométrica

Es como el de Little y Lodish, pero incluyendo otro coeficiente:

$$D_{ij} = \beta \cdot \left[\frac{A_i \cdot A_j}{P} \right]^\alpha \Leftrightarrow$$

Para estimarlo se han vuelto a tomar logaritmos en ambos miembros, llegándose a la siguiente expresión:

$$\ln(D_{ij}) = \ln \beta + \alpha \cdot \ln \left[\frac{A_i \cdot A_j}{P} \right]$$

Se trata de estimar una ecuación lineal que, a diferencia de la anterior, tiene un término independiente $\ln(\beta)$. Los resultados aparecen en la **Tabla 2.41**:

α	1,008206587	Error estándar α	0,006089318
$\ln \beta$	0,404256287	Error estándar $\ln \beta$	0,041728312
r^2	0,770874721	Error estándar var. ind.	1,458988843
F	27.413,33147	Grados de libertad	8.148
Suma de los cuadrados de la regresión	58.353,34539	Suma de los cuadrados de los residuos	17.344,22752

Tabla 2.41

Así queda definida la ecuación, sabiendo que si $\ln(\beta) \approx 0,4043$, entonces $\beta \approx 1,4982$:

$$D_{ij} = 1,4982 \cdot \left[\frac{A_i \cdot A_j}{P} \right]^{1,0082}$$

El coeficiente de determinación es ligeramente mejor, pero los errores estándar asociados a los parámetros α y β son mayores. Además plantea los mismos problemas que el modelo original de Lodish. Por tanto, lo rechazamos al no suponer ninguna mejora significativa.

6.3.2. Ecuación logarítmica sin término independiente

Tendría una expresión como sigue:

$$D_{ij} = \alpha \cdot \ln \left[\frac{A_i \cdot A_j}{P} \right]$$

Los resultados de la regresión, que se muestran en la **Tabla 2.42**, indican la poca explicación del modelo. El coeficiente de determinación r^2 apenas llega al 9%, y presenta errores muy elevados. Se debe rechazar, por tanto, este modelo.

α	2.164,144477	Error estándar α	55,76382255
r^2	0,090696845	Error estándar var. ind.	34.497,96998
F	812,8076798	Grados de libertad	8.149
Suma de los cuadrados de la regresión	9,6733E+11	Suma de los cuadrados de los residuos	9,69821E+12

Tabla 2.42

6.3.3. Ecuación logarítmica con término independiente

$$D_{ij} = \beta + \alpha \cdot \ln \left[\frac{A_i \cdot A_j}{P} \right]$$

Este modelo explica mejor la nube de puntos, pero todavía el coeficiente de determinación es muy pequeño: $r^2=15\%$ (véase la **Tabla 2.43**). Se ha aumentado la significación y se han reducido los errores, pero siguen siendo muy elevados.

α	5.402,771871	Error estándar α	138,6330218
β	-24.072,0536	Error estándar β	950,0115572
r^2	0,157114859	Error estándar var. ind.	33.216,20703
F	1.518,797526	Grados de libertad	8.148
Suma de los cuadrados de la regresión	1,67571E+12	Suma de los cuadrados de los residuos	8,98982E+12

Tabla 2.43

6.3.4. Ecuación exponencial

$$D_{ij} = \beta \cdot e^{\alpha \cdot (A_i \cdot A_j) / P}$$

Para hacer la regresión lineal hemos transformado esa expresión de la siguiente manera:

$$\ln(D_{ij}) = \ln \beta + \alpha \cdot \frac{A_i \cdot A_j}{P} \cdot \ln e = \ln \beta + \alpha \cdot \frac{A_i \cdot A_j}{P}$$

El valor estimado del coeficiente $\ln \beta$ ha sido de 6,4689, luego β valdrá 647,77 y el modelo se formulará así:

$$D_{ij} = 644,77 \cdot e^{0,000055 \cdot (A_i \cdot A_j) / P}$$

La regresión alcanza un coeficiente de determinación del 13% (Tabla 2.44), de modo que, aunque el modelo sea significativo en su conjunto, está explicando muy poco el fenómeno de la duplicación.

α	5,54483E-05	Error estándar α	1,58947E-06
$\ln \beta$	6,468894	Error estándar $\ln \beta$	0,032684565
r^2	0,129947378	Error estándar var. ind.	2,843074417
F	1.216,950801	Grados de libertad	8.148
Suma de los cuadrados de la regresión	9.836,701112	Suma de los cuadrados de los residuos	65.860,8718

Tabla 2.44

6.3.5. Ecuación lineal sin término independiente

Es la forma más sencilla que puede tomar la ecuación:

$$D_{ij} = \alpha \cdot \frac{A_i \cdot A_j}{P}$$

Si α valiese 1, se estaría cumpliendo estrictamente la hipótesis de duplicación libre o aleatoria. Sin embargo, vemos en la Tabla 2.45 que la mejor estimación lleva a que $\alpha=1,7036$. Comprobamos que mejora el coeficiente de determinación del modelo de

Little y Lodish: $r^2=86\%$. Por otra parte, ahora la suma de los cuadrados de la regresión y de los residuos es bastante mayor.

α	1,70357209	Error estándar α	0,00723146
r^2	0,86051397	Error estándar var. ind.	13.424,41985
F	50.982,0779	Grados de libertad	8.264
Suma de los cuadrados de la regresión	9,1877E+12	Suma de los cuadrados de los residuos	1,4893E+12

Tabla 2.45

6.3.6. Ecuación lineal con término independiente

Vamos a buscar una ecuación que no pase por el cruce de ejes, y por tanto tenga esta expresión.

$$D_{ij} = \beta + \alpha \cdot \frac{A_i \cdot A_j}{p}$$

Apenas se ha encontrado mejora en el modelo (véase la **Tabla 2.46**). Los valores de α y β estimados son muy parecidos a los del modelo anterior. Por tanto, dadas las pocas diferencias entre los modelos lineales vistos, es aconsejable escoger el que no tiene término independiente.

α	1,69407163	Error estándar α	730,0756364
β	0,00749126	Error estándar β	152,9688394
r^2	0,86089744	Error estándar var. ind.	13.406,76554
F	51.139,2135	Grados de libertad	8.263
Suma de los cuadrados de la regresión	9,1918E+12	Suma de los cuadrados de los residuos	1,4852E+12

Tabla 2.46

6.4. Conclusiones

Hemos observado que el modelo de Little y Lodish explicaba el 76% de la regresión correspondiente. El coeficiente r^2 se ha conseguido incrementar hasta el 86%, mediante otra ecuación lineal. Tales valores indican un buen ajuste, quizá suficientes en la explicación de otros fenómenos. Sin embargo, las duplicaciones deben ser estimadas con mayor precisión, debido a su utilidad posterior: pretendemos que sean el instrumento adecuado para estimar el alcance y la distribución de contactos y, por tanto, hay que evitar que los errores iniciales se puedan multiplicar en los resultados finales. En los modelos estocásticos veremos planteamientos alternativos, que intentan estimar las duplicaciones a partir de distintas hipótesis.

7. Modelo de Lozano

Lozano⁷⁹, antiguo Director-Gerente del EGM, observó una estrecha relación entre las audiencias del último periodo y los hábitos de audiencia, en los datos del EGM (I Parte, §B). Bajo la hipótesis de que son iguales, llegó a formular el alcance neto de un conjunto de soportes en función de su audiencia bruta (A), del nº de inserciones (n) y del tamaño del target (P), con la ventaja de no necesitar de las duplicaciones:

$$C = \frac{A}{\frac{A}{P} + \frac{2}{3} - \frac{1}{3} \cdot \log \frac{100A}{nP}}$$

Entiéndase que sólo se hace una inserción por soporte. Lozano testó la fórmula en 500 casos, sobre 15 targets distintos, en Diarios, Suplementos, Revistas, Radio y Televisión. Comentaba que "el resultado ha sido positivo, hallando la respuesta buscada con diferencias mínimas". Sin embargo, no aparecen los resultados ni más explicación de las campañas, así que no se pueden verificar tales afirmaciones. También concluye que las estimaciones son mejores cuando las audiencias de los soportes son mayores (por ejemplo, en Televisión).

Al testarlo en nuestras campañas de revistas, he observado una sobrestimación media del alcance neto igual al 43'53%, con una desviación típica del 20'61%. Los errores disminuyen cuando intervienen más soportes, pero son siempre inaceptables. Encuentro tres explicaciones ante esta grave equivocación:

1. Los parámetros de la ecuación se han estimado a partir de los valores extremos, en vez de usar la regresión u otras técnicas;
2. La forma de la ecuación no tiene fundamentación teórica; y
3. Los tests se han podido realizar sobre un gran número de soportes.

⁷⁹ LOZANO CHOZA, Domingo. "Probabilidad de la audiencia de los medios de comunicación". *Investigación y Marketing*. Nº 19. Noviembre 1985. Págs. 23-6.

B) MODELOS ESTOCÁSTICOS

Conceptos previos e hipótesis fundamentales

Un modelo estocástico⁸⁰ es un sistema de hipótesis acerca del comportamiento de los lectores entre un grupo de individuos hacia un conjunto de soportes. La premisa fundamental es que la exposición a los soportes es un fenómeno aleatorio sujeto a las específicas leyes de la probabilidad. Un modelo completo debe contener al menos tres hipótesis:

a) Si los individuos pueden o no considerarse como idénticos con respecto a su probabilidad de exposición a un soporte dado. Se habla de *homogeneidad de los individuos* cuando todos los componentes del grupo tienen la misma probabilidad de exposición frente a un soporte, y de *heterogeneidad de los individuos* cuando existe una cierta distribución de las probabilidades individuales de exposición a un soporte.

b) Si los soportes pueden o no considerarse como idénticos en lo referente a sus probabilidades de exposición. Existe *homogeneidad de los soportes* cuando la probabilidad de exposición de un individuo no varía con el soporte considerado, mientras que la *heterogeneidad de soportes* se da cuando esa probabilidad no es constante.

c) Si los soportes pueden o no considerarse como independientes con respecto a los demás. Tendremos *independencia de audiencias* cuando la exposición de un individuo a los soportes constituye un conjunto de sucesos mutuamente independientes, de forma que la exposición a un soporte no modifica la probabilidad de exposición a otro soporte. La *dependencia de las audiencias* se da cuando la exposición a un soporte no es aleatoria, sino que está condicionado por otra u otras exposiciones.

⁸⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 342.

Como cada una de estas hipótesis tiene dos alternativas, teóricamente habría 8 (2³) familias de modelos de audiencia neta. Sin embargo, sólo seis de ellos tienen sentido porque si se admite la heterogeneidad de los individuos no se puede mantener la independencia de las audiencias. Los modelos estocásticos se clasifican según estas hipótesis implícitas; a continuación se recogen primero los que mantienen la homogeneidad de individuos (Tabla 2.47) y después la heterogeneidad de los individuos (Tabla 2.48).

	Homogeneidad de soportes	Heterogeneidad de soportes
Independencia de audiencias	Binomial	Sainsbury
Dependencia de audiencias	Kwerel	Sainsbury modificado Goodhardt-Ehrenberg Headen- Klomp.-Rust Rust-Zimmer-Leone Agregación Secuencial Independencia Condicional Hofmans ampliado

Tabla 2.47: Modelos estocásticos con **homogeneidad de los individuos**.
Fuente: Chandon (op. cit.)

	Homogeneidad de soportes	Heterogeneidad de soportes
Dependencia de audiencias	Metherringham	Matriz Beta

Tabla 2.48: Modelos estocásticos con **heterogeneidad de los individuos**.
Fuente: Chandon (op. cit.)

En las páginas siguientes se expone estos modelos, aplicándose a los datos de nuestra ya citada investigación, de manera que los resultados nos sirvan para medir la precisión de los modelos y para testar las hipótesis implícitas. El modelo Matriz Beta, el menos restrictivo en sus hipótesis, se explica en la **IV Parte**.

1. Modelo binomial

1.1. Definición de hipótesis y funcionamiento

El binomial es el modelo estocástico más restrictivo en cuanto a sus hipótesis⁸¹:

- Homogeneidad de los individuos: todos los individuos tienen la misma probabilidad de contactar con un soporte i .

- Homogeneidad de los soportes: la audiencia de todos los soportes es la misma:

$$A_1 = A_2 = A_3 = \dots = A_n$$

- Independencia de las audiencias: se cumple para todo soporte i, j, k, \dots, n :

$$P(S_i / S_j) = P(S_i)$$

$$P(S_i \cap S_j) = P(S_i) \cdot P(S_j)$$

$$P(S_i \cap S_j \cap S_k) = P(S_i) \cdot P(S_j) \cdot P(S_k)$$

...

$$P(S_i \cap S_2 \cap \dots \cap S_n) = P(S_1) \cdot P(S_2) \cdot \dots \cdot P(S_n)$$

Como consecuencia de las dos primeras hipótesis la probabilidad de exposición (p) es la misma para todos los individuos y todos los soportes. Siendo \bar{A} audiencia media para n soportes:

$$\bar{A} = \frac{\sum_{i=1}^n A_i}{n}$$

La probabilidad de exposición p de un individuo cualquiera a cualquiera de los soportes será: $p = \bar{A} / P$. Por la hipótesis de la independencia de audiencias, la exposición a un soporte es una variable de Bernoulli. Por tanto la probabilidad de que un individuo esté expuesto exactamente a k entre n soportes:

⁸¹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 345-7.

$$P(x = k / n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

y la probabilidad de que esté expuesto al menos a k soportes será:

$$P(x \geq k / n, p) = \sum_{x=k}^n \binom{n}{x} p^x (1 - p)^{n-x}$$

Para el caso especial de $k=1$, o la probabilidad de que un individuo esté expuesto al menos a un soporte:

$$P(x \geq 1 / n, p) = 1 - P(x = 0 / n, p) = 1 - (1 - p)^n$$

Como esa probabilidad es igual para todos los individuos, el alcance neto (C_n) resulta de multiplicar por la población:

$$C_n = [1 - (1 - p)^n] \cdot P$$

Chandon⁸² estimó el alcance neto de cinco revistas para nueve segmentos de población, a partir de los datos de "1964 Politz Magazine Study". El modelo binomial producía sobrestimaciones del 17,7% de media, permaneciendo estables estos errores entre los *target*. Al considerar once revistas, la sobrestimación media aumentaba hasta el 23%. Como los errores permanecen constantes entre los segmentos de población, Chandon sugiere que se podría establecer un conjunto de deflatores, según el número de revistas, para eliminar esas sobrestimaciones sistemáticas.

⁸² CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 348-50.

1.2. Ejemplo

La ventaja de este modelo es que estima el alcance neto a partir únicamente de las audiencias individuales de los soportes. Recordemos en la **Tabla 2.49** los datos de nuestra combinación de revistas:

REVISTA	Audiencia
Muy Interesante	309.918
Teleprograma	236.902
El Jueves	130.265
Teleindiscreta	129.992

Tabla 2.49

Como los soportes se suponen homogéneos hallamos la audiencia media (\bar{A}) de las cuatro revistas:

$$\bar{A} = \frac{\sum_{i=1}^n A_i}{n} = \frac{309.918 + 236.902 + 130.265 + 129.992}{4} \approx 201.769$$

Todos los individuos tendrán la misma probabilidad (p) de contactar con alguno de esos soportes:

$$p = \frac{\bar{A}}{P} = \frac{201.769}{1.455.001} \approx 0'1387$$

Y, suponiendo que es independiente contactar con un soporte o con otro, podemos hallar el alcance neto de las cuatro revistas:

$$C_n = [1 - (1-p)^n] \cdot P = [1 - (1-0'1387)^4] \cdot 1.455.001 \approx 654.179$$

La distribución de contactos se calcula como sigue:

$$E_0^4 = 1.455.001 \cdot \binom{4}{0} 0'1387^0 (1-0'1387)^4 \approx 800.822 \text{ (contactan con 0 revistas)}$$

$$E_1^4 = 1.455.001 \cdot \binom{4}{1} 0'1387^1 (1 - 0'1387)^3 \approx 515.726 \text{ (contactan con 1 de las revistas)}$$

$$E_2^4 = 1.455.001 \cdot \binom{4}{2} 0'1387^2 (1 - 0'1387)^2 \approx 124.547 \text{ (contactan con 2 de las revistas)}$$

$$E_3^4 = 1.455.001 \cdot \binom{4}{3} 0'1387^3 (1 - 0'1387)^1 \approx 13.368 \text{ (contactan con 3 de las revistas)}$$

$$E_4^4 = 1.455.001 \cdot \binom{4}{4} 0'1387^4 (1 - 0'1387)^0 \approx 538 \text{ (contactan con las 4 revistas)}$$

1.3. Tests del alcance neto estimado

El modelo binomial ha producido sobrestimaciones del 15,4% de media, una cifra que puede considerarse bastante elevada, teniendo además en cuenta que la mayoría de las estimaciones cometen errores entre el 8,8% y el 22%. Como los errores en valor absoluto tienen la misma dimensión que sin valor absoluto, en ningún caso se producen subestimaciones. Podría emplearse un factor de corrección, teniendo en cuenta la tendencia de los errores. No se observa relación alguna entre el público objetivo y el comportamiento de los errores (**Tabla 2.50**), de manera que las variaciones son aleatorias. En cuanto a los criterios de selección de soportes se ha observado que no influyen en los errores, y no se producen diferencias mínimamente significativas.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
H, 14-34, A-MA	11,7992	4,3361	11,7992	4,3361
H, 14-34, MM	9,9867	4,3765	9,9867	4,3765
H, 14-34, MB-B	10,9753	4,7418	10,9753	4,7418
H, 35-54, A-MA	16,0816	5,0992	16,0816	5,0992
H, 35-54, MM	13,6786	4,2103	13,6786	4,2103
H, 35-54, MB-B	16,7472	5,2831	16,7472	5,2831
H, >54, A-MA	15,8287	5,8438	15,8287	5,8438
H, >54, MM	13,8226	4,2562	13,8226	4,2562
H, >54, MB-B	17,9156	5,9463	17,9156	5,9463
M, 14-34, A-MA	11,2951	5,1831	11,2951	5,1831
M, 14-34, MM	10,2320	2,9838	10,2320	2,9838
M, 14-34, MB-B	10,2280	1,8656	10,2280	1,8656
M, 35-54, A-MA	17,0770	11,5027	17,0770	11,5027
M, 35-54, MM	11,9621	3,4426	11,9621	3,4426
M, 35-54, MB-B	17,4359	5,3995	17,4359	5,3995
M, >54, A-MA	17,7875	3,7851	17,7875	3,7851
M, >54, MM	20,5301	2,6465	20,5301	2,6465
M, >54, MB-B	26,8408	6,0064	26,8408	6,0064
TOTAL	19,1565	5,0727	19,1565	5,0727
TODOS LOS CASOS	15,4454	6,6085	15,4454	6,6085

Tabla 2.50: Errores del modelo binomial, según el público objetivo.

La estimación del alcance neto empeora cuando la campaña va incluyendo más soportes. Desde un error del 8,6% con 4 revistas, se incrementa progresivamente hasta el 19% con 20 revistas (Tabla 2.51). La desviación estándar, que mide la dispersión respecto a la media, se mantiene aproximadamente constante -salvo la última alternativa- deduciéndose que, aunque aumente la dimensión de los errores, éstos se mantienen igualmente agrupados.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
4	8,6053	4,7972	8,6053	4,7972
8	14,2354	5,2787	14,2354	5,2787
12	16,0840	4,9261	16,0840	4,9261
16	17,4380	5,0210	17,4380	5,0210
20	18,9823	7,3177	18,9823	7,3177
TODOS LOS CASOS	15,4454	6,6085	15,4454	6,6085

Tabla 2.51: Errores del modelo binomial, según el número de soportes.

1.4. Contrastes de la distribución de contactos

Se han estimado por el Modelo Binomial las distribuciones de contactos de 375 campañas, seleccionadas entre las 402 habituales (para simplificar, sólo las de 4, 8, 12, 16 y 20 inserciones). Se comparan después con las ofrecidas por el TOM2. Ya habíamos comentado los errores cometidos por el TOM en la distribución de contactos (**I Parte, §D.3**): ahora vamos a analizar las diferencias entre estimaciones, ambas sujetas a error.

En cada campaña se halla la diferencia total, en valores simples o absolutos, $\sum_{i=1}^n (f_i^B - f_i^T)$ ó $\sum_{i=1}^n |f_i^B - f_i^T|$, siendo f_i^B la frecuencia del contacto i estimada por el Binomial y f_i^T la frecuencia del contacto i estimada por el TOM. Dichas funciones, que están limitadas entre 0 y 1, expresan las diferencias de frecuencias en la parte contactada de la población y, por tanto, están limitadas por la cobertura de la campaña.

Las diferencias totales de las 375 campañas se han analizado a partir de las medias de los grupos característicos. En conjunto las diferencias son importantes, ya que por término medio alcanzan el 6'73% ó el 10'38% en valores absolutos (**Tabla 2.52**). Decíamos antes que el modelo binomial sobrestimaba el alcance, y ahora observamos que sus frecuencias son habitualmente mayores que las del TOM. Las variaciones entre *targets* no presentan tendencias definidas, salvo una difusa superioridad en las diferencias absolutas de los hombres sobre las mujeres.

POBLACIÓN OBJETIVO	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
H, 14-34, A-MA	6,48%	2,97%	9,97%	4,83%	20
H, 14-34, MM	5,53%	2,52%	8,55%	4,24%	20
H, 14-34, MB-B	5,61%	2,64%	8,64%	4,28%	20
H, 35-54, A-MA	7,09%	3,29%	10,83%	4,87%	20
H, 35-54, MM	6,26%	2,86%	9,48%	4,04%	20
H, 35-54, MB-B	5,39%	2,10%	8,89%	3,03%	20
H, >54, A-MA	6,27%	2,98%	10,13%	4,76%	20
H, >54, MM	5,46%	2,19%	8,50%	3,15%	20
H, >54, MB-B	3,88%	1,61%	6,96%	2,40%	20
M, 14-34, A-MA	6,79%	3,75%	11,12%	6,51%	20
M, 14-34, MM	6,19%	2,32%	10,17%	3,79%	20
M, 14-34, MB-B	6,35%	1,87%	10,71%	3,21%	20
M, 35-54, A-MA	7,81%	2,45%	12,89%	4,10%	20
M, 35-54, MM	7,02%	2,53%	10,61%	4,06%	20
M, 35-54, MB-B	7,79%	3,09%	11,52%	4,21%	20
M, >54, A-MA	9,32%	2,98%	15,10%	4,93%	20
M, >54, MM	9,89%	2,53%	15,65%	4,03%	20
M, >54, MB-B	7,17%	2,34%	11,62%	2,99%	20
TOTAL	7,92%	3,36%	4,51%	2,52%	15
TODOS LOS CASOS	6,73%	2,97%	10,38%	4,71%	375

Tabla 2.52: Contrastes entre las distribuciones estimadas por los modelos Binomial y TOM (según el público objetivo).

Las diferencias son muy similares aunque las campañas se hayan seleccionado según distintos criterios (ver **Tabla 2.53**). Las desviaciones estándar ocupan aproximadamente la mitad de la media, igual que en todos los casos, porque esta agrupación no se ajusta realmente a la diversidad.

CRITERIO	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
rentabilidad	6,45%	3,06%	9,91%	4,88%	95
cobertura	7,11%	2,89%	10,83%	4,74%	95
rentabilidad+cobertura	6,88%	2,95%	10,54%	4,74%	95
afinidad	6,48%	2,96%	10,25%	4,48%	90
TODOS LOS CASOS	6,73%	2,97%	10,38%	4,71%	375

Tabla 2.53: Contrastes entre las distribuciones estimadas por los modelos Binomial y TOM (según el criterio de selección de soportes).

El número de soportes sí que está determinando las diferencias entre el Binomial y el TOM (Tabla 2.54). Tanto en valores simples como en absolutos, las diferencias aumentan sustancialmente con los soportes. Los grupos resultantes son más homogéneos que antes. La interpretación es sencilla: con pocas inserciones el TOM producía grandes errores, que ahora se asemejan a los del Binomial; según aumentaban las inserciones el TOM iba mejorando, pero ahora no sucede así con el Binomial; en conclusión, en campañas de pocas inserciones el TOM estima casi tan mal como el Binomial, que tiene las hipótesis menos realistas.

NÚMERO DE SOPORTES	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
4	2,48%	1,43%	4,11%	2,56%	75
8	5,69%	1,83%	8,74%	3,33%	75
12	7,44%	1,69%	11,36%	3,10%	75
16	8,65%	1,64%	13,28%	2,89%	75
20	9,39%	1,67%	14,43%	2,77%	75
TODOS LOS CASOS	6,73%	2,97%	10,38%	4,71%	375

Tabla 2.54: Contrastes entre las distribuciones estimadas por los modelos Binomial y TOM (según el nº de soportes).

2. Modelo de Sainsbury

2.1. Hipótesis y descripción

De origen británico, se publica por primera vez en el artículo referenciado⁸³, pero su mayor difusión comienza cuando fue recogido por Caffyn y Sagovsky⁸⁴. Se trata de un modelo que considera el fenómeno de la exposición de soportes como un proceso aleatorio. Al igual que el modelo binomial, permite estimar el alcance neto a partir de las audiencias de los soportes, sin que sea necesario conocer las duplicaciones. Como veremos posteriormente, las duplicaciones sí son necesarias en el modelo de Sainsbury modificado. Antes de desarrollar el modelo, expondremos las hipótesis de partida:

- Hipótesis de *homogeneidad de los individuos* de la población objetivo: todas las personas de la población objetivo tienen la misma probabilidad de exposición frente a un soporte dado. Igual que en el modelo binomial.

- Hipótesis de *heterogeneidad de soportes*: la probabilidad de exposición que tiene un individuo varía con el soporte considerado. Al contrario que el modelo binomial:

$$A_1 \neq A_2 \neq A_3 \neq \dots \neq A_n \neq A$$

- Hipótesis de *independencia de audiencias*: la exposición a los soportes está formada por conjuntos mutuamente excluyentes. La audiencia sigue siendo una variable Bernouilli.

La probabilidad de exposición $P(S_i)$ de un individuo a un soporte, se define matemáticamente de la siguiente manera:

$$P(S_i) = \frac{A_i}{P} \quad 0 \leq P(S_i) \leq 1$$

⁸³ LEE, A. M. y BURKART, A. J.: "Some Optimization Problems in Advertising Media Planning". *Operational Research Quarterly*. Vol. 11. Nº 3. Septiembre 1960.

⁸⁴ CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods". *Journal of Advertising Research*. Marzo 1963. Vol 3. Nº1. Págs. 22-3.

Supongamos ahora tres soportes S_1 , S_2 y S_3 . Un individuo cualquiera puede pertenecer a uno a varios a la vez. Se trata pues de sucesos probabilísticos que pueden ser simultáneos⁸⁵. A partir del cálculo de probabilidades, se dice que la probabilidad de que un individuo esté expuesto al menos a uno de esos soportes es la siguiente:

$$\begin{aligned} P(S_1 \cup S_2 \cup S_3) &= P(S_1) + P(S_2) + P(S_3) \\ &\quad - P(S_1 \cap S_2) - P(S_1 \cap S_3) - P(S_2 \cap S_3) \\ &\quad + P(S_1 \cap S_2 \cap S_3) \end{aligned}$$

Como se ha establecido la hipótesis de independencia de esos sucesos aleatorios, se puede transformar:

$$\begin{aligned} P(S_1 \cup S_2 \cup S_3) &= P(S_1) + P(S_2) + P(S_3) \\ &\quad - P(S_1) \cdot P(S_2) - P(S_1) \cdot P(S_3) - P(S_2) \cdot P(S_3) \\ &\quad + P(S_1) \cdot P(S_2) \cdot P(S_3) \end{aligned}$$

Restando la unidad y cambiando el signo en ambos miembros, y sacando factor común en el segundo miembro, se obtiene:

$$1 - P(S_1 \cup S_2 \cup S_3) = [1 - P(S_1)] \cdot [1 - P(S_2)] \cdot [1 - P(S_3)]$$

Tal ecuación se puede generalizar para el caso de n soportes:

$$1 - P(S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n) = [1 - P(S_1)] \cdot [1 - P(S_2)] \cdot [1 - P(S_3)] \cdot \dots \cdot [1 - P(S_n)]$$

También se puede determinar a partir de los complementarios de tres soportes, denotando $P(\bar{S}_i)$ como el complementario de $P(S_i)$:

$$P(S_1 \cup S_2 \cup S_3) = 1 - P(\bar{S}_1 \cap \bar{S}_2 \cap \bar{S}_3) = 1 - P(\bar{S}_1) \cdot P(\bar{S}_2) \cdot P(\bar{S}_3)$$

El alcance neto se determina finalmente por la siguiente expresión:

$$C_n = P(S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n) \cdot P$$

⁸⁵ MARTÍNEZ RAMOS, EMILIO. "El uso de los medios de comunicación en marketing y publicidad". Ed. Akal/Comunicación. Madrid, 1992. Pág. 54.

2.2. Ejemplo

En la **Tabla 2.55** se determina la probabilidad de exposición de un individuo cualquiera ante la revista i [$P(S_i)$], o su probabilidad complementaria, es decir, la de no exposición [$P(\bar{S}_i)$]. El público objetivo (P) estaba compuesto por 1.455.001 personas.

REVISTA	Audiencias	$P(S_i) = \frac{A_i}{P}$	$P(\bar{S}_i) = \frac{P - A_i}{P}$
Muy Interesante	309.918	0,213	0,787
Teleprograma	236.902	0,1628	0,8372
El Jueves	130.265	0,0895	0,9105
Teleindiscreta	129.992	0,0893	0,9107

Tabla 2.55

Con los datos anteriores, se calcula la probabilidad de que un individuo cualquiera del público objetivo contacte al menos con una de las cuatro revistas:

$$1 - P(S_1 \cup S_2 \cup S_3 \cup S_4) = [1 - P(S_1)] \cdot [1 - P(S_2)] \cdot [1 - P(S_3)] \cdot [1 - P(S_4)] \Leftrightarrow$$

$$1 - P(S_1 \cup S_2 \cup S_3 \cup S_4) = [1 - 0,213] \cdot [1 - 0,1628] \cdot [1 - 0,0895] \cdot [1 - 0,0893] \Leftrightarrow$$

$$1 - P(S_1 \cup S_2 \cup S_3 \cup S_4) \approx 0,5462 \Leftrightarrow P(S_1 \cup S_2 \cup S_3 \cup S_4) \approx 0,4538$$

A través de los complementarios, se habría procedido así:

$$\begin{aligned} P(S_1 \cup S_2 \cup S_3 \cup S_4) &= 1 - P(\bar{S}_1 \cap \bar{S}_2 \cap \bar{S}_3 \cap \bar{S}_4) = 1 - P(\bar{S}_1) \cdot P(\bar{S}_2) \cdot P(\bar{S}_3) \cdot P(\bar{S}_4) = \\ &= 1 - 0,787 \cdot 0,8372 \cdot 0,9105 \cdot 0,9107 \approx 0,4538 \end{aligned}$$

Finalmente se calcula el alcance neto de las cuatro revistas:

$$C_4 = P(S_1 \cup S_2 \cup S_3 \cup S_4) \cdot P = 0,4538 \cdot 1.455.001 \approx 660.311$$

2.3. Aplicación al alcance neto de nuestras campañas

Chandon, al testar este modelo en los citados datos de Politz, encontró unas sobrestimaciones medias del 18,3% para 5 revistas y del 24,6% para 11 revistas. En nuestro caso hemos encontrado una sobrestimación media del 16,3%, con una desviación típica respecto a la media del 6,2%. Estos errores son ligeramente mayores que en el anterior modelo estocástico, el binomial. A diferencia de todos los modelos anteriores, se observa que los errores de los *targets* siguen una tendencia explicable: se hacen mayores al aumentar la edad de referencia (Tabla 2.56); sin embargo, no hay relación entre los errores y el sexo o la clase social. La correlación entre los errores y la edad refuta, de manera indirecta, la hipótesis de homogeneidad de los individuos, ya que la precisión de las estimaciones depende de las características de los individuos.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	12,8103	4,1990	12,8103	4,1990
H, 14-34, MM	11,1309	4,4258	11,1309	4,4258
H, 14-34, MB-B	12,3443	4,5297	12,3443	4,5297
H, 35-54, A-MA	16,4460	5,2323	16,4460	5,2323
H, 35-54, MM	14,7146	4,3089	14,7146	4,3089
H, 35-54, MB-B	17,9951	5,6240	17,9951	5,6240
H, >54, A-MA	16,2085	5,9929	16,2085	5,9929
H, >54, MM	14,7814	4,4131	14,7814	4,4131
H, >54, MB-B	18,8330	6,1352	18,8330	6,1352
M, 14-34, A-MA	11,9059	5,2108	11,9059	5,2108
M, 14-34, MM	11,1427	2,8843	11,1427	2,8843
M, 14-34, MB-B	11,6656	1,8933	11,6656	1,8933
M, 35-54, A-MA	15,1348	2,6849	15,1348	2,6849
M, 35-54, MM	12,6884	3,6409	12,6884	3,6409
M, 35-54, MB-B	18,7084	5,5866	18,7084	5,5866
M, >54, A-MA	19,1698	3,6753	19,1698	3,6753
M, >54, MM	21,4866	2,9914	21,4866	2,9914
M, >54, MB-B	27,6690	6,3225	27,6690	6,3225
TOTAL	20,1065	5,1833	20,1065	5,1833
TODOS LOS CASOS	16,2716	6,1914	16,2716	6,1914

Tabla 2.56: Errores del modelo de Sainsbury, según el público objetivo.

En la **Tabla 2.57**, aparecen los errores de estimación del alcance neto, según el criterio de selección de soportes. No se observan variaciones significativas entre las distintas alternativas, ni en la media ni la desviación estándar, por lo se deduce que no influye el modo de seleccionar las revistas.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
rentabilidad	15,6385	6,5188	15,6385	6,5188
cobertura	16,4598	6,3448	16,4598	6,3448
rentabilidad+cobertura	16,1071	6,4525	16,1071	6,4525
afinidad	16,9757	5,2649	16,9757	5,2649
TODOS LOS CASOS	16,2716	6,1914	16,2716	6,1914

Tabla 2.57: Errores del modelo de Sainsbury, según el criterio de selección de soportes.

La estimación por el método de Sainsbury ha resultado más imperfecta a medida que las campañas incluyen más soportes (**Tabla 2.58**). Las sobrestimaciones crecen progresivamente del 9,2% para 4 revistas al 19,4% para 20 revistas. La desviación estándar, sin embargo, permanece casi constante.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
4	9,2092	4,5616	9,2092	4,5616
8	15,1446	5,0926	15,1446	5,0926
12	17,1093	4,8446	17,1093	4,8446
16	18,4875	5,0504	18,4875	5,0504
20	19,4607	5,1499	19,4607	5,1499
TODOS LOS CASOS	16,2716	6,1914	16,2716	6,1914

Tabla 2.58: Errores del modelo de Sainsbury, según el número de soportes.

3. Modelo de Sainsbury modificado

3.1. Modificación y nuevo desarrollo

Este modelo⁸⁶ adopta una hipótesis distinta al anterior: en vez de considerar las audiencias independientes, supone que son interdependientes. Resumimos aquí las hipótesis del nuevo modelo:

- a) Homogeneidad de los individuos.
- b) Heterogeneidad de los soportes.
- c) Interdependencia de las audiencias, es decir, la exposición a los soportes está formada por conjuntos dependientes entre sí. Dicho de otra forma, la probabilidad de exposición a un soporte está condicionando la probabilidad de exposición a otro.

Como antes, la probabilidad de ser alcanzado, es la diferencia entre uno y la probabilidad de la intersección de los complementarios:

$$P(S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n) = 1 - P(\bar{S}_1 \cap \bar{S}_2 \cap \bar{S}_3 \cap \dots \cap \bar{S}_n)$$

Pero al aceptar la hipótesis de interdependencia de las audiencias:

$$P(\bar{S}_1 \cap \bar{S}_2 \cap \bar{S}_3 \cap \dots \cap \bar{S}_n) = P(\bar{S}_1) \cdot P(\bar{S}_2 / \bar{S}_1) \cdot P(\bar{S}_3 / \bar{S}_1 \cap \bar{S}_2) \cdot \dots \cdot P(\bar{S}_n / \bar{S}_1 \cap \bar{S}_2 \cap \dots \cap \bar{S}_{n-1})$$

La anterior expresión supondría conocer todas las intersecciones entre soportes; pero, teniendo en cuenta las hipótesis iniciales del modelo, se pueden obtener estas simplificaciones:

$$P(\bar{S}_3 / \bar{S}_1 \cap \bar{S}_2) = P(\bar{S}_3 / \bar{S}_1)$$

$$P(\bar{S}_4 / \bar{S}_1 \cap \bar{S}_2 \cap \bar{S}_3) = P(\bar{S}_4 / \bar{S}_1)$$

.....

$$P(\bar{S}_n / \bar{S}_1 \cap \bar{S}_2 \cap \dots \cap \bar{S}_{n-1}) = P(\bar{S}_n / \bar{S}_1)$$

⁸⁶ DÍEZ DE CASTRO, ENRIQUE CARLOS y MARTÍN ARMARIO, ENRIQUE. "Planificación Publicitaria". Ed. Pirámide. Madrid, 1993. Págs. 285-6.

Sustituyendo en las ecuaciones anteriores, obtenemos:

$$P(S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n) = 1 - P(\bar{S}_1) \cdot P(\bar{S}_2 / \bar{S}_1) \cdot P(\bar{S}_3 / \bar{S}_1) \dots P(\bar{S}_n / \bar{S}_1)$$

Por el teorema de Bayes se sabe que:

$$P(A \cap B) = P(A) \cdot P(B / A) = P(B) \cdot P(A / B)$$

$$P(A / B) = \frac{P(A \cap B)}{P(B)} ; \quad P(\bar{A} / \bar{B}) = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{B})}$$

Sustituyendo en la expresión anterior:

$$P(S_1 \cup S_2 \cup S_3 \cup \dots \cup S_n) = 1 - P(\bar{S}_1) \frac{P(\bar{S}_1 \cap \bar{S}_2)}{P(\bar{S}_1)} \frac{P(\bar{S}_1 \cap \bar{S}_3)}{P(\bar{S}_1)} \dots \frac{P(\bar{S}_1 \cap \bar{S}_n)}{P(\bar{S}_1)}$$

En cada caso, $P(\bar{S}_1 \cap \bar{S}_i)$ se calcula así:

$$P(S_1 \cup S_i) = P(S_1) + P(S_i) - P(S_1 \cap S_i)$$

$$P(S_1 \cup S_i) = 1 - P(\bar{S}_1 \cap \bar{S}_i)$$

$$P(\bar{S}_1 \cap \bar{S}_i) = 1 - [P(S_1) + P(S_i) - P(S_1 \cap S_i)]$$

Como se puede ver, se ha utilizado el soporte S_1 como pivote para las transformaciones del modelo de Sainsbury. Se podrían haber utilizado otros soportes, obteniéndose un resultado distinto en cada caso. Entonces, ¿cuál debería ser el soporte que ajustara la audiencia no cubierta de los otros soportes? Caffyn y Sagovsky⁸⁷ señalaban que debía ser aquel soporte con más audiencia; sin embargo, si la pareja con mayor duplicación no incluye al soporte de mayor audiencia, se debería elegir el que tenga más audiencia de esa pareja con mayor duplicación. En nuestros 402 casos hemos seleccionado como soporte pivote el de mayor audiencia, en vez del de mayor duplicación, para facilitar el cálculo; no obstante, son pocos los casos donde el soporte

⁸⁷ CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods". *Journal of Advertising Research*. Marzo 1963. Vol 3. Nº1. Págs. 22-3.

de mayor audiencia no tiene la mayor duplicación y los resultados finales no presentan una variación significativa.

3.2. Ejemplo

Vamos a considerar la revista Muy Interesante como soporte pivote. Por eso recogemos en la **Tabla 2.59** sus duplicaciones con las otras revistas, así como la probabilidad de que un individuo de la población objetivo esté en contacto con una de esas duplicaciones.

REVISTAS (S_i)	Audiencias (A_i)	Duplicaciones con Muy Interesante (D_{ij})	$P(S_i) = \frac{A_i}{P}$	$P(S_i \cap S_j) = \frac{D_{ij}}{P}$
Muy Interesan.	309.918	-	0,213	-
Teleprograma	236.902	55.208	0,1628	0,0379
El Jueves	130.265	47.445	0,0895	0,0326
Teleindiscreta	129.992	30.036	0,0893	0,0206

Tabla 2.59

Necesitamos hallar la probabilidad de un individuo no contacte simultáneamente con Muy Interesante y una de las otras revistas:

$$P(\bar{S}_1 \cap \bar{S}_2) = 1 - [P(S_1) + P(S_2) - P(S_1 \cap S_2)] = 1 - [0,213 + 0,1628 - 0,0379] \approx 0,662$$

$$P(\bar{S}_1 \cap \bar{S}_3) = 1 - [P(S_1) + P(S_3) - P(S_1 \cap S_3)] = 1 - [0,213 + 0,0895 - 0,0326] \approx 0,73$$

$$P(\bar{S}_1 \cap \bar{S}_4) = 1 - [P(S_1) + P(S_4) - P(S_1 \cap S_4)] = 1 - [0,213 + 0,0893 - 0,0206] \approx 0,718$$

Ya estamos en condiciones de determinar la probabilidad de contactar al menos con una de las revistas y, a continuación, conocer el alcance neto:

$$\begin{aligned}
 P(S_1 \cup S_2 \cup S_3 \cup S_4) &= 1 - P(\bar{S}_1) \cdot \frac{P(\bar{S}_1 \cap \bar{S}_2)}{P(\bar{S}_1)} \cdot \frac{P(\bar{S}_1 \cap \bar{S}_3)}{P(\bar{S}_1)} \cdot \frac{P(\bar{S}_1 \cap \bar{S}_4)}{P(\bar{S}_1)} = \\
 &= 1 - 0,787 \cdot \frac{0,662}{0,787} \cdot \frac{0,73}{0,787} \cdot \frac{0,718}{0,787} \approx 0,4395
 \end{aligned}$$

$$C_4 = P(S_1 \cup S_2 \cup S_3 \cup S_4) \cdot P = 0,4395 \cdot 1.455.001 \approx 639.500$$

3.3. Aplicación al alcance neto de nuestras campañas

La modificación en el modelo de Sainsbury mejora los resultados, pero dista todavía bastante de los modelos ad-hoc. Chandon había obtenido una sobrestimación media del 6,2% para 5 revistas y del 12,7% para 11 revistas. En nuestro caso, la sobrestimación media se sitúa ahora en el 10% con una desviación típica del $\pm 6\%$. En todos los casos se sobrestima el alcance neto y, por ello, no varían los errores en valor absoluto. Comparando los errores según los públicos objetivo (Tabla 2.60), observamos que los errores aumentan cuando el intervalo de edad es mayor, igual que el modelo simplificado; pero ahora también vemos que los hombres tienen menores errores que las mujeres; respecto a la clase social, parece que no tiene relación. Se vuelve a refutar la hipótesis de homogeneidad de los individuos, mantenida en este modelo por Sainsbury. Aunque no se recoge aquí, la estimación no varía con el criterio de selección de soportes.

POBLACIÓN OBJETIVO	MEDIA e%	DES. EST. e%	MEDIA e%	DES. EST. e%
H, 14-34, A-MA	7,7194	3,7363	7,7194	3,7363
H, 14-34, MM	8,0497	3,6252	8,0497	3,6252
H, 14-34, MB-B	7,4714	3,5273	7,4714	3,5273
H, 35-54, A-MA	8,1559	5,6322	8,4807	5,1024
H, 35-54, MM	10,6156	4,2476	10,6156	4,2476
H, 35-54, MB-B	12,1411	4,8602	12,1411	4,8602
H, >54, A-MA	11,4658	5,8581	11,4658	5,8581
H, >54, MM	13,7766	14,1745	13,7766	14,1745
H, >54, MB-B	12,9803	5,3851	12,9803	5,3851
M, 14-34, A-MA	6,6144	3,9788	6,6436	3,9273
M, 14-34, MM	5,4494	2,6427	5,4703	2,5969
M, 14-34, MB-B	4,9688	2,0144	4,9688	2,0144
M, 35-54, A-MA	6,3145	2,3006	6,3145	2,3006
M, 35-54, MM	7,6751	3,1949	7,6751	3,1949
M, 35-54, MB-B	10,8394	4,5890	10,8394	4,5890
M, >54, A-MA	9,2576	4,2130	9,2576	4,2130
M, >54, MM	12,5730	3,9330	12,5730	3,9330
M, >54, MB-B	14,5414	5,1037	14,5414	5,1037
TOTAL	14,8916	4,9310	14,8916	4,9310
TODOS LOS CASOS	10,0439	6,0296	10,0625	5,9984

Tabla 2.60: Errores del modelo de Sainsbury modificado, según el público objetivo.

Nuevamente la estimación empeora con mayor número de soportes, como refleja la **Tabla 2.61**. Hay que señalar que la modificación del modelo de Sainsbury mejora sustancialmente las estimaciones, pero sobre todo con pocos vehículos: ahora el error medio cometido con 4 revistas es del 3,6%, casi la tercera parte del 9,2% obtenido con el modelo simplificado. Pero si estamos ante una campaña con muchos soportes, hemos de desechar este modelo, pues los errores son relevantes; no obstante, dado que la desviación estándar es relativamente pequeña, podría emplearse un factor que corrigiera la sobrestimación media.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	3,5730	3,9450	3,5863	3,9327
8	9,0050	7,7781	9,0050	7,7781
12	10,6258	3,5660	10,6258	3,5660
16	11,9104	4,1296	11,9970	3,8672
20	12,8165	3,8467	12,8165	3,8467
TODOS LOS CASOS	10,0439	6,0296	10,0625	5,9984

Tabla 2.61: Errores del modelo de Sainsbury modificado, según el nº de soportes.

4. Modelo de Kwerel

4.1. Hipótesis y razonamientos

Kwerel estableció un modelo para estimar la audiencia no duplicada de una combinación de soportes a partir de la suma de audiencias individuales y de la suma de la duplicación entre los pares⁸⁸. Se basaba en el llamado principio de máximo condicional de entropía⁸⁹, rigurosamente validado y aplicable a la exposición de audiencia de una combinación de soportes. Hace las siguientes hipótesis:

a) homogeneidad de individuos,

b) homogeneidad de soportes,

c) interdependencia de audiencias: las audiencias constituyen un conjunto de sucesos mutuamente excluyentes; más específicamente, si una persona ha estado expuesta a un soporte, se modifica la probabilidad de que esa persona esté expuesta a otros soportes; en otras palabras, la probabilidad condicional de exponerse a un soporte difiere de su probabilidad no condicional. Analíticamente se puede expresar:

$$P(S_i / S_j) \neq P(S_i)$$

$$P(S_i \cap S_j) \neq P(S_i) \cdot P(S_j / S_i) = P(S_j) \cdot P(S_i / S_j)$$

$$P(S_i \cap S_j \cap S_k) \neq P(S_i) \cdot P(S_j \cap S_k / S_i)$$

...

$$P(S_1 \cap S_2 \cap \dots \cap S_n) \neq P(S_1) \cdot P(S_1 \cap S_2 \cap \dots \cap S_{n-1} / S_1)$$

El alcance neto de los n soportes estará comprendida entre los siguientes dos límites:

$$L_{\min} = (h+1) \cdot \bar{A} - \frac{h \cdot (h+1)}{2} \cdot \bar{D} \qquad L_{\max} = A - \frac{2}{n} \cdot D$$

⁸⁸ KWEREL, SEYMOUR M. "Estimating Unduplicated Audience and Exposure Distribution". *Journal of Advertising Research*. Vol. 9. N° 2. Junio de 1969. Págs. 46-53.

⁸⁹ KWEREL, SEYMOUR M. "Estimating the Unduplicated Audience of a Combination of Media Vehicles - Integrated Theory and Estimation Method". Ph. D. Dissertation. Columbia University. 1964.

definiéndose A y D , por la homogeneidad de individuos, como la media de las audiencias y la media de las duplicaciones respectivamente:

$$\bar{A} = \frac{A}{n} ; \quad \bar{D} = \frac{D}{\frac{n(n-1)}{2}}$$

y, h como el número mínimo entre las dos cantidades:

$$h = \min [(n-1) , k]$$

siendo k la parte entera del cociente $\frac{\bar{A}}{\bar{D}}$.

El alcance neto se estima entonces mediante la fórmula:

$$C = \frac{2 \cdot L_{\max} \cdot L_{\min}}{L_{\max} + L_{\min}}$$

Kwerel observó que cuando el "ratio crítico" (definido como el cociente D/A) se acercaba al valor de $\frac{n-1}{4}$, y cuanto mayor fuese n , los límites tienden a hacerse relativamente más amplios. El efecto es que la estimación del alcance puede ser más imprecisa.

4.2. Ejemplo

Recordemos en la **Tabla 2.62** las audiencias (en negrita) y las duplicaciones de las cuatro revistas del ejemplo habitual:

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.62

Como parte de que los individuos y los soportes son homogéneos, no interesan los datos de audiencia individuales, sino la suma total y la media de las audiencias brutas y de las duplicaciones:

$$A = \sum_{i=1}^4 A_i = 309.918 + 236.902 + 130.265 + 129.992 = 807.077$$

$$D = 55.208 + 47.445 + 30.036 + 23.935 + 29.574 + 11.555 = 197.753$$

$$\bar{A} = \frac{A}{n} = \frac{807.077}{4} \approx 201.769$$

$$\bar{D} = \frac{D}{\frac{n(n-1)}{2}} = \frac{197.753}{\frac{4(4-1)}{2}} \approx 32.959$$

$$\frac{\bar{A}}{\bar{D}} = \frac{201.769}{32.959} \approx 6,12. \text{ Luego } k, \text{ el valor entero de este cociente, vale } 6.$$

h es valor mínimo entre $(n-1=3)$ y $(k=6)$. Por tanto, $h=3$.

Se determinan los límites que acotan el alcance neto:

$$L_{\max} = A - \frac{2}{n} \cdot D = 807.077 - \frac{2}{4} \cdot 197.753 \approx 708.200$$

$$L_{\min} = (h+1) \cdot \bar{A} - \frac{h \cdot (h+1)}{2} \cdot \bar{D} = (3+1) \cdot 201.769 - \frac{3 \cdot (3+1)}{2} \cdot 32.959 \approx 609.322$$

y se estima dicho alcance:

$$C = \frac{2 \cdot L_{\max} \cdot L_{\min}}{L_{\max} + L_{\min}} = \frac{2 \cdot 708.200 \cdot 609.322}{708.200 + 609.322} \approx 654.975$$

4.3. Tests del alcance neto

Evaluando el alcance neto según Kwerel, hemos obtenido en las 402 campañas una subestimación media del 7%. Chandon había encontrado una subestimación media del 4,2% para 5 revistas y 13,5% para 11 revistas. En nuestro caso, obviando el signo del error, para no eliminar sobrestimaciones con subestimaciones, se calcula que el error medio era del 8,3%, por ahora el mejor de los modelos estocásticos. El cruce entre los errores y la población objetivo presenta variaciones entre las distintas alternativas (**Tabla 2.63**): la media de errores para los hombres se acerca al 5%, mientras que para las mujeres es del 8%; parece también que al subir la edad, el error es mayor; también resulta significativo que si no segmentamos la población, el error es del 12% (este valor se ha de tomar con más precaución pues se basa en 42 casos). Estos análisis contradicen la hipótesis de homogeneidad de los individuos mantenida en este modelo.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
H, 14-34, A-MA	-6,4889	7,3831	7,9249	5,7205
H, 14-34, MM	-6,5765	6,9539	7,6835	5,6338
H, 14-34, MB-B	-5,3630	6,7642	6,9050	5,0835
H, 35-54, A-MA	-6,2531	7,4636	7,7777	5,7608
H, 35-54, MM	-4,8317	7,1538	6,9451	4,9979
H, 35-54, MB-B	-2,4920	5,9944	5,2599	3,6533
H, >54, A-MA	-5,2563	5,6988	6,3625	4,3528
H, >54, MM	-7,6420	6,8033	8,5045	5,6237
H, >54, MB-B	0,6628	2,0865	1,8478	1,1053
M, 14-34, A-MA	-7,4955	8,1116	9,3314	5,7688
M, 14-34, MM	-8,6301	7,8823	9,8444	6,2061
M, 14-34, MB-B	-9,2002	7,3771	9,9778	6,2229
M, 35-54, A-MA	-9,9117	8,1960	10,7167	7,0494
M, 35-54, MM	-8,8640	6,6839	9,8784	4,9661
M, 35-54, MB-B	-7,8031	7,4666	9,1848	5,5715
M, >54, A-MA	-6,9995	4,8014	7,5365	3,8518
M, >54, MM	-8,8707	5,8335	9,4800	4,7181
M, >54, MB-B	-4,9527	4,5028	6,1074	2,6138
TOTAL	-11,7044	8,0908	12,6299	6,5108
TODOS LOS CASOS	-7,0422	7,2755	8,3478	5,7262

Tabla 2.63: Errores del modelo de Kwerel, según el público objetivo.

Al analizar separadamente los errores según el criterio empleado para seleccionar los soportes (**Tabla 2.64**) hemos encontrado pequeñas diferencias, quizá irrelevantes dada la magnitud de los errores. Es el primer modelo, de los hasta ahora recogidos, que presenta ajustes diferentes según los objetivos de medios: rentabilidad, cobertura o afinidad.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-6,5643	7,3298	7,9891	5,7262
cobertura	-8,3539	7,2574	9,3375	5,9245
rentabilidad+cobertura	-7,6365	7,2980	8,7863	5,8476
afinidad	-5,3918	6,9484	7,1120	5,1507
TODOS LOS CASOS	-7,0422	7,2755	8,3478	5,7262

Tabla 2.64: Errores del modelo de Kwerel, según el criterio de selección de soportes.

La estimación por el modelo de Kwerel empeora sustancialmente cuando consideramos un mayor número de soportes (**Tabla 2.65**). Con 4 vehículos se produce una sobrestimación del 2,4%, con 8 vehículos pasa a ser una subestimación del 1,9%, creciendo aceleradamente la subestimación hasta los 20 soportes. Nos encontramos con otro modelo que parece ser formulado para pocas inserciones: a partir de 8, los errores son bastante representativos, aunque su desviación estándar permanece bastante constante.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	2,4480	1,0822	2,5179	0,9051
8	-1,9158	4,0450	3,6537	2,5597
12	-7,6403	3,7544	7,7571	3,5033
16	-11,5641	3,7520	11,5641	3,7520
20	-14,0550	3,9338	14,0550	3,9338
TODOS LOS CASOS	-7,0422	7,2755	8,3478	5,7262

Tabla 2.65: Errores del modelo de Kwerel, según el número de soportes.

5. Modelos de duplicaciones *entre-soportes*

5.1. MODELO DE GOODHARDT Y EHRENBURG

Goodhardt y Ehrenberg⁹⁰ propusieron un modelo en el cual las duplicaciones observadas en una serie de vehículos no es aleatoria. Lo validaron para el Medio Televisión, pero al tener una formulación bastante general se puede aplicar también a las Revistas. Establecieron la ley de la duplicación de los telespectadores: "el número de individuos que ven simultáneamente dos programas de TV en días diferentes de la misma semana puede determinarse simplemente a partir de la audiencia de los dos programas, sin tener en cuenta otros factores, como el contenido del programa".

Plantean la investigación comparando programas dentro de un mismo canal y entre dos canales de TV que existían en ese momento en Gran Bretaña. La muestra se compone por amas de casa londinenses, pero los resultados son similares a los conseguidos con otras muestras. La ecuación fundamental relaciona la audiencia duplicada con las audiencias de los soportes, corregidas por la población, y con la constante k :

$$A_{ij} = k \cdot \left[\frac{A_i \cdot A_j}{P} \right] \quad \forall i < n, j < n; i \neq j$$

que se puede expresar en términos de probabilidad:

$$p_{ij} = k \cdot p_i \cdot p_j$$

siendo k la media de los k_{ij} obtenidos en cada par de vehículos:

$$k = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij}}{\binom{n}{2}} \quad k_{ij} = \frac{p_{ij}}{p_i \cdot p_j}$$

⁹⁰ GOODHARDT, Gerald J. and EHRENBURG, A.S.C.: "Duplication of Television Viewing Between and Within Channels". *Journal of Marketing Research*. Mayo de 1969. Págs. 169-78.

Demostraron que los valores de k_{ij} podrían sustituirse por la constante k , sin que se produjera un error significativo; k aproximadamente valdría 1 si la duplicación de programas era de canales de TV distintos, y 1'5 si era del mismo canal. Un mayor valor de k supone una duplicación superior, quizá explicable por la fidelidad al canal por parte de los telespectadores. Sin embargo, no influye el contenido del programa ni la hora en que se emita.

Señalan varias limitaciones: habría que estudiar una desviación sistemática detectada en dos franjas horarias; también han aparecido desviaciones irregulares que no se han podido explicar; hay programas dentro de una misma serie que contarán con una duplicación mayor que lo habitual; y, por último, las propias limitaciones de la muestra seleccionada.

Aunque parece que el contenido del programa debiera influir en el tamaño de la duplicación, señalan⁹¹ que no se ha encontrado relación entre los televidentes y las categorías de los programas (noticias, documentales, etc).

Los autores no se refieren a combinaciones que contengan más de dos inserciones. No obstante recogemos aquí ese artículo por la relevancia que tiene el estudio de las duplicaciones entre soportes.

Vamos a testar esta hipótesis de duplicación constante. Para ello empleamos los datos habituales. Tenemos las duplicaciones para 19 públicos y para 30 revistas: en total, 8.265 duplicaciones observadas $[19 \cdot 30 \cdot (30-1)/2]$.

Hemos cogido primeramente el total de la población. Tenemos 435 duplicaciones, y despejamos los valores de k_{ij} como se había explicado antes. A simple vista, parece claro que k_{ij} no es constante: los valores no se concentran en torno a su media $\bar{k}=2'48847$. Pero midamos su dispersión a través de la desviación estándar:

⁹¹ EHRENBURG, A.S.C. and GOODHARDT, Gerald J.: "Practical Applications of the Duplication of Viewing Law". *Journal of Marketing Research Society*, 11. Enero de 1969. Págs. 6-24.

$$S = \sqrt{\frac{\sum_{i=1}^{435} (k_{ij} - \bar{k})^2}{435}} = 2,266$$

Por otra parte el valor mínimo de k_{ij} es de 0'55 y el máximo de 23'301. Se observa pues que la distribución tiene un recorrido muy amplio que refuta la hipótesis de la constancia de k_{ij} . Además el coeficiente de variación de Pearson⁹² nos indica la dispersión respecto a la media es bastante grande, ya que la desviación típica supone más del noventa por ciento de la media:

$$V = \frac{S}{\bar{k}} = \frac{2,266}{2,488} \cdot 100 = 91,1\%$$

Dados estos resultados, no parece necesario analizar todas las duplicaciones, pero al menos sí nos vamos a referir al primero de los segmentos (hombres, de 14 a 34 años y de clase social alta y media alta): \bar{k} valdría 2'835, oscilando k_{ij} entre un mínimo de 0'037 y un máximo de 94'219; la desviación típica sería 6'25 y el coeficiente de variación de Pearson de 220'5%. Por tanto, es insostenible mantener que k_{ij} es constante.

La hipótesis de constancia de k también había sido rechazada en la referida investigación de Chandon, para las 55 duplicaciones de 11 revistas del Politz Magazine Study.

Ejemplo

Pero antes de pasar a otro modelo queremos referirnos nuevamente a la campaña nº 10 (cfr. **RTDOS.WK1**), para ejemplificar los planteamientos teóricos que se han recogido aquí; en la **Tabla 2.66** aparecen las seis posibles duplicaciones de los soportes Muy Interesante (S_1), Teleprograma (S_2), El Jueves (S_3) y Teleindiscreta (S_4).

⁹² MARTÍN-GUZMÁN, M. PILAR y MARTÍN PLIEGO, F. JAVIER. "Curso Básico de Estadística Económica". Editorial AC. 3ª edición. 1989. Págs. 70-1.

S_i	A_i	$p_i = \frac{A_i}{P}$	S_j	A_j	$p_j = \frac{A_j}{P}$	D_{ij}	$p_{ij} = \frac{D_{ij}}{P}$	$k_{ij} = \frac{p_{ij}}{p_i \cdot p_j}$
S_1	309.918	0,213	S_2	236.902	0,163	55.208	0,038	1,094
S_1	309.918	0,213	S_3	130.265	0,090	47.445	0,033	1,710
S_1	309.918	0,213	S_4	129.992	0,089	30.036	0,021	1,085
S_2	236.902	0,163	S_3	130.265	0,090	23.935	0,016	1,128
S_2	236.902	0,163	S_4	129.992	0,089	29.574	0,020	1,397
S_3	130.265	0,090	S_4	129.992	0,089	11.555	0,008	0,993

Tabla 2.66

Como se ve en la última columna, no se puede afirmar que k_{ij} permanezca constante entre los distintos pares de soportes.

5.2. MODELO DE HEADEN-KLOMPMAKER-RUST

Headen, Klompmaker y Rust⁹³ investigaron la duplicación entre programas de Televisión en Estados Unidos, con un planteamiento más completo que el de Goodhardt y Ehrenberg en Gran Bretaña. Incorporan tres variables: tipo de programa, horario y exposición repetida (del mismo programa). La estimación de la duplicación (p_{ij}) se basaba en un modelo obtenido por regresión:

$$p_{ij} = b_0 (b_1^{x_1})(b_2^{x_2})(b_3^{x_3})(b_4^{x_4})(p_i p_j)^{b_5} e^u$$

siendo

b_0, b_1, \dots, b_5 = coeficientes de regresión;

x_1 = 0 si los programas pertenecen a diferentes canales y 1 si pertenecen al mismo

x_2 = 0 si son tipos de programas diferentes y 1 si son iguales

x_3 = 0 si están en diferentes partes del día y 1 si están en la misma

x_4 = 0 cuando son diferentes programas y 1 cuando son diferentes episodios del mismo programa

u = perturbación aleatoria.

Los programas se clasificaron en diez tipos: drama, drama-acción, drama psicológico, juego, charla o variedad, películas, noticias, comedia, deportes y otros. El día se dividía en cinco franjas horarias: 7 a.m.-10 a.m.; 10 a.m.-5 p.m.; 5 p.m.-8 p.m.; 8 p.m.-11 p.m.; y 11 p.m.-después.

Aunque algunas variables presentaban covarianzas ligeramente altas, el modelo no tenía problemas importantes de multicolinealidad. Se aplicó una regresión de tipo *stepwise* para determinar el orden de prioridad de las variables: 1º las probabilidades de exposición; 2º el canal; 3º el tipo de programa; 4º la parte del día; y 5º la exposición

93 HEADEN, Robert S.; KLOMPMAKER, Jay E.; y RUST, Roland T. "The Duplication of Viewing Law and Television Media Schedule Evaluation". *Journal of Marketing Research*. Vol. 16. Agosto 1979. Págs. 333-40.

repetida. Gracias a las tres últimas, el R^2 se incrementaba del 0'72 al 0'82: esto es una buena muestra de la mejoría sobre el modelo Goodhardt-Ehrenberg. Además, todas las variables eran significativas, lo mismo que el modelo en su conjunto. Además este modelo permite estimar tanto la duplicación *intra*-programa como la duplicación *entre*-programas.

Los autores indican que las duplicaciones son útiles para seleccionar y evaluar campañas alternativas pero, sobre todo, pueden servir para estimar la distribución de contactos; así lo muestran en una sencilla campaña, cuya estimación es mejor con las duplicaciones de H-K-R que con las de G-E, usando el modelo Beta Binomial. Sin duda el propósito es muy útil, como ya hemos comentado varias veces; pero es necesario que las duplicaciones sean muy precisas pues, de lo contrario, los errores de la distribución de contactos pueden ser significativos. No se trata simplemente de explicar el fenómeno de la duplicación, sino de estimar fielmente las duplicaciones en cada caso.

5.3. MODELO DE RUST-ZIMMER-LEONE

Rust, Zimmer y Leone⁹⁴ desarrollaron un nuevo modelo para estimar la duplicación *intra*-programas, añadiendo el tipo de programa y el horario como variables explicativas (obsérvese que es igual al modelo de H-K-R pero sin los productorios):

$$p_{ii} = b_0 (p_i^2)^{b_1} \prod_i (c_i^{x_i}) \prod_j (d_j^{y_j})$$

siendo:

$x_i = 1$ si pertenece al tipo de programa i y 0 su contrario

$y_j = 1$ si el programa es de la franja horaria j y 0 su contrario

b_0, b_1, c_i, d_j = coeficientes de regresión.

Las estimaciones resultaban más precisas que las de H-K-R, ya se tratara de duplicaciones en la misma semana o en distintas semanas. Y bastante más precisas que las del modelo de Rice-Leckenby⁹⁵, que tiene esta forma (el coeficiente k se ajusta a la audiencia y tipo de programa):

$$p_{ii} = k \cdot (p_i)^2$$

Por otra parte, Cannon⁹⁶ hizo un modelo para estimar la duplicación de audiencia en segmentos, a partir de la duplicación en el mercado total; se basó en la confrontación empírica de 20 pares de revistas en 8 *targets*, con datos del Target Group Index (1977); llegó a esta expresión, para dos soportes i y j (en mayúsculas la duplicación y las audiencias sobre la población total, y en minúsculas sobre el *target*):

⁹⁴ RUST, Roland T.; ZIMMER, Mary R.; y LEONE, Robert P. "Estimating the Duplicated Audience of Media Vehicles in National Advertising Schedules". *Journal of Advertising*. Vol 15. Nº 3. 1986. Págs. 30-7.

⁹⁵ RICE, Marshall D. y LECKENBY, John D. "Predicting Within-Vehicle Television Duplication". *Proceedings of the American Academy of Advertising*. 1984.

⁹⁶ CANNON, Hugh M. "Reach and Frequency Estimates for Specialized Target Markets". *Journal of Advertising Research*. Vol 23. Nº 3. Junio/Julio 1983. Págs. 45-50.

$$d_{ij} = \frac{D_{ij} \cdot (a_i + a_j)}{A_i + A_j}$$

Se trata de determinar la intersección de tres conjuntos: los que contactan con el soporte i , los que contactan con el soporte j y los que pertenecen al *target*; se asimila a una triplicación (**Figura 2.9**). Rust, Zimmer y Leone la calculan como la media de las tres posibles duplicaciones, que suponen aleatorias:

$$d_{ij} = \frac{1}{3} \cdot \left[\frac{D_{ij} \cdot a_i}{A_i} + \frac{D_{ij} \cdot a_j}{A_j} + \frac{a_i \cdot a_j}{P} \right]$$

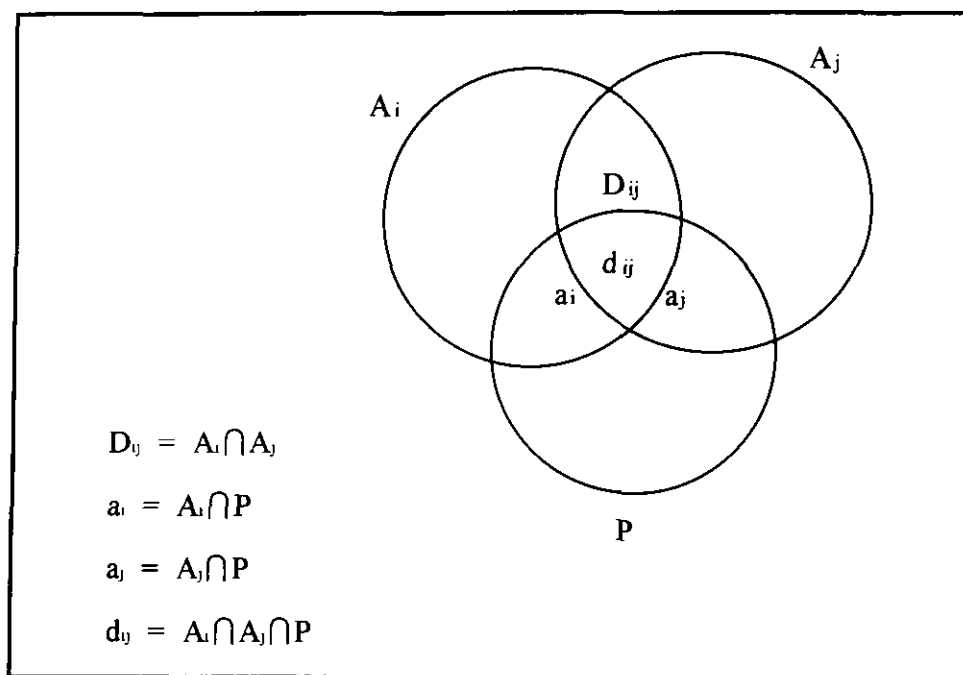


Figura 2.9

R-Z-L testaron esta fórmula con los datos Simmons⁹⁷ (1977), sobre las respuestas diarias de 5.652 individuos, evaluando 1.210 duplicaciones entre: (1) distintos programas de TV, (2) un programa en la misma semana, (3) un programa en diferentes semanas; (5) distintas revistas, (6) dos números de la misma revista, (7) duplicación

⁹⁷ SIMMONS MEDIA STUDIES. "Selective Markets and the Media Reaching Them". New York. 1978.

revista-programa. Todo referido a diez *targets*, que combinaban el sexo y cinco intervalos de edad. Los resultados indicaban una mejoría del modelo de Cannon en los casos (5) y (7). Pero insisto en lo dicho anteriormente: calcular la distribución de contactos a partir de duplicaciones estimadas, exige que éstas sean todavía más precisas.

6. Modelos de agregación secuencial

6.1. Desarrollo e hipótesis implícitas

Estos modelos⁹⁸ parten de varias hipótesis: las probabilidades de exposición son homogéneas para cada individuo y heterogéneas para cada soporte. Además, según la duplicación sea o no aleatoria, tendremos modelos de agregación con duplicación variable o constante. Estiman el alcance de dos soportes y, una vez realizado, los consideran como un único soporte, cuya audiencia duplicará con la de un tercer soporte que ha sido agregado, obteniéndose el alcance de esa combinación de soportes. Se procede de esta manera hasta incorporar todos los soportes de la campaña.

Para determinar el alcance de tres revistas, suponiendo que la duplicación es constante e igual a k , primero se calcula el alcance de dos revistas:

$$A_{1,2} = A_1 + A_2 - k \cdot \frac{A_1 \cdot A_2}{P}$$

Agregamos el tercer soporte, cuya audiencia duplica con el alcance de los dos soportes anteriores:

$$A_{1,2,3} = A_{1,2} + A_3 - k \cdot \frac{A_{1,2} \cdot A_3}{P}$$

Remplazando $A_{1,2}$ en la ecuación anterior obtenemos:

$$A_{1,2,3} = A_1 + A_2 - k \cdot \frac{A_1 \cdot A_2}{P} + A_3 - k \cdot \frac{\left[A_1 + A_2 - k \cdot \frac{A_1 \cdot A_2}{P} \right] \cdot A_3}{P} \Rightarrow$$

$$A_{1,2,3} = A_1 + A_2 + A_3 - \frac{k}{P} \cdot \left[A_1 \cdot A_2 + \left(A_1 + A_2 - k \cdot \frac{A_1 \cdot A_2}{P} \right) \cdot A_3 \right] \Rightarrow$$

⁹⁸ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 388-91.

$$A_{1,2,3} = A_1 + A_2 + A_3 - \frac{k}{p} \cdot [A_1 \cdot A_2 + A_1 \cdot A_3 + A_2 \cdot A_3] + \frac{k^2}{p^2} \cdot A_1 \cdot A_2 \cdot A_3$$

De la ecuación anterior se deduce que no importa el orden de agregación de los soportes: $A_{1,2,3} = A_{1,3,2} = A_{2,3,1} = A_{3,2,1}$.

Bajo la hipótesis de que las duplicaciones, triplicaciones,..., son aleatorias, podríamos definir la suma de duplicaciones (S_2), la suma de triplicaciones (S_3),..., hasta la suma de las duplicaciones de orden n (S_n):

$$S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_i \cdot A_j}{p}$$

$$S_3 = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n \frac{A_i \cdot A_j \cdot A_k}{p^2}$$

...

$$S_n = \sum_{i=1}^1 \sum_{j=i+1}^2 \dots \sum_{y=x+1}^{n-1} \sum_{z=y+1}^n \frac{A_i \cdot A_j \cdot \dots \cdot A_y \cdot A_z}{p^{n-1}} = \frac{A_i \cdot A_j \cdot \dots \cdot A_y \cdot A_z}{p^{n-1}}$$

De las fórmulas anteriores y para el caso de los tres soportes, se deduce que la audiencia neta ($A_{1,2,3}$) se puede determinar restando de la audiencia bruta (A) las duplicaciones multiplicadas por k y sumando las triplicaciones multiplicadas por k^2 :

$$A_{1,2,3} = A - k \cdot S_2 + k^2 \cdot S_3$$

Y también se podría generalizar para el caso de n soportes:

$$A_{1,2,3,\dots,n} = A - k \cdot S_2 + k^2 \cdot S_3 - \dots \pm k^{n-1} \cdot S_n \quad [1]$$

Sabiendo que el alcance neto se define como el alcance bruto menos las duplicaciones (D), más las triplicaciones (T), menos las cuadruplicaciones (C),..., y así hasta las n repeticiones:

$$A_{1,2,3,\dots,n} = A - D + T - C + \dots \pm N \quad [2]$$

Identificando los términos de la ecuación [1] con los de la ecuación [2], encontramos que los dos términos son idénticos bajo la hipótesis de duplicación constante. Además, la comparación entre los términos restantes define las hipótesis implícitas sobre el método de agregación secuencial a emplear. Primeramente vamos a demostrar la igualdad entre los segundos términos, la referente a las duplicaciones:

$$k \cdot S_2 = D \quad \Leftrightarrow \quad k = \frac{D}{S_2}$$

Supongamos que k es la media de las k_{ij} resultantes de los pares posibles de soportes:

$$k = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij}}{\binom{n}{2}} \quad \Rightarrow \quad k = P \cdot \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_{ij}}{A_i \cdot A_j}}{\binom{n}{2}}$$

Por otra parte, D/S_2 viene determinado por la siguiente expresión:

$$\frac{D}{S_2} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_i \cdot A_j}{P}}$$

Se comprueba que no se puede igualar k y D/S_2 , después de haber despejado P :

$$\frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_{ij}}{A_i \cdot A_j}}{\binom{n}{2}} \neq \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n A_i \cdot A_j}$$

Los dos miembros de la ecuación son iguales sólo si todos los ratios $\frac{A_{ij}}{A_i \cdot A_j}$ son iguales, que es precisamente la hipótesis de duplicación constante. Cuanto mayor sea la varianza de $\binom{n}{2}$, mayor será el error de la estimación.

Si no admitimos la hipótesis de duplicación constante, el parámetro constante k lo tendríamos que transformar en el coeficiente variable k' :

$$k' = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_i \cdot A_j}{P}} = \frac{\text{suma de las duplicaciones observadas}}{\text{suma de las duplicaciones aleatorias}}$$

El valor de k' dependerá de las revistas que compongan la combinación. Este coeficiente será una medida de la correlación dentro de las n revistas: $k'=1$ cuando la exposición es aleatoria; si $k'>1$, existe correlación positiva; si $k'<1$, existe correlación negativa. Finalmente, k' se reduce a la constante de Goodhardt-Ehrenberg cuando se consideran sólo dos soportes.

Dada la dificultad de cálculo de este método, no lo hemos aplicado a nuestra investigación. Se recomienda emplear habitualmente modelos más sencillos para estimar el alcance neto de una combinación de soportes. Además los errores alcanzados por Chandon en su pequeña muestra son elevados, sobre todo cuando crece el número de soportes: para 11 revistas, la subestimación media es del 29% con duplicación constante y del 19% con duplicación variable.

6.2. Ejemplo

Recordemos en la **Tabla 2.67** los datos del ejemplo al que nos venimos refiriendo: las audiencias (en negrilla) y las duplicaciones de las cuatro revistas:

REVISTA	Muy Interes.	Teleprograma	El Jueves	Teleindiscreta
Muy Interes.	309.918	55.208	47.445	30.036
Teleprograma		236.902	23.935	29.574
El Jueves			130.265	11.555
Teleindiscreta				129.992

Tabla 2.67

Como en la práctica no se cumple la hipótesis de duplicación constante, podemos demostrar que $k \neq D/S_2$:

$$k = P \cdot \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_{ij}}{A_i \cdot A_j}}{\binom{n}{2}} = \frac{1.455.002}{6} \cdot \left(\frac{55.208}{309.918 \cdot 236.902} + \frac{47.445}{309.918 \cdot 130.265} + \right. \\ \left. + \frac{30.036}{309.918 \cdot 129.992} + \frac{23.935}{236.902 \cdot 130.265} + \frac{29.574}{236.902 \cdot 129.992} + \frac{11.555}{130.265 \cdot 129.992} \right) \approx 1,2346$$

$$D = \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij} = 55.208 + 47.445 + 30.036 + 23.935 + 29.574 + 11.555 = 197.753$$

$$S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{A_i \cdot A_j}{P} = \frac{309.918 \cdot 236.902}{1.455.002} + \frac{309.918 \cdot 130.265}{1.455.002} + \frac{309.918 \cdot 129.992}{1.455.002} + \\ + \frac{236.902 \cdot 130.265}{1.455.002} + \frac{236.902 \cdot 129.992}{1.455.002} + \frac{130.265 \cdot 129.992}{1.455.002} \approx 159.909$$

$$k = 1,2346 \neq \frac{D}{S_2} = \frac{197.753}{159.909} = 1,2367. \text{ (como queríamos demostrar)}$$

Supongamos ahora que la **duplicación es constante**; para esas cuatro revistas, k toma el valor 1'2346; vamos a estimar ahora el alcance neto según la agregación secuencial. Primero necesitamos obtener la suma de triplicaciones (S_3) y de cuadruplicaciones (S_4):

$$S_3 = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n \frac{A_i \cdot A_j \cdot A_k}{P^2} = \frac{309.918 \cdot 236.902 \cdot 130.265}{1.455.002^2} + \frac{309.918 \cdot 236.902 \cdot 129.992}{1.455.002^2} +$$

$$+ \frac{309.918 \cdot 130.265 \cdot 129.992}{1.455.002^2} + \frac{236.902 \cdot 130.265 \cdot 129.992}{1.455.002^2} \approx 13.400$$

$$S_4 = \sum_{i=1}^1 \sum_{j=i+1}^2 \sum_{k=j+1}^3 \sum_{l=k+1}^4 \frac{A_i \cdot A_j \cdot A_k \cdot A_l}{P^{4-1}} = \frac{A_1 \cdot A_2 \cdot A_3 \cdot A_4}{P^3} =$$

$$= \frac{309.918 \cdot 236.902 \cdot 130.265 \cdot 129.992}{1.455.002^3} \approx 404$$

Con estos datos ya podemos hallar la audiencia neta de las cuatro revistas:

$$C_4 = A - k \cdot S_2 + k^2 \cdot S_3 - k^3 \cdot S_4 =$$

$$= 807.077 - 1,2346 \cdot 159.909 + 1,2346^2 \cdot 13.400 - 1,2346^3 \cdot 404 \approx 629.318$$

Bajo la hipótesis de **duplicación no constante**, se tomaría el valor de $k'=D/S_2=1'2367$. Este coeficiente nos indica que existe correlación positiva ($k'>1$) entre las audiencias de las cuatro revistas. Para estimar el alcance neto procedemos de la misma forma que antes, pero sustituyendo k por k' :

$$C_4 = A - (k') \cdot S_2 + (k')^2 \cdot S_3 - (k')^3 \cdot S_4 =$$

$$= 807.077 - 1,2367 \cdot 159.909 + 1,2367^2 \cdot 13.400 - 1,2367^3 \cdot 404 \approx 629.048$$

7. Modelo de independencia condicional

7.1. Hipótesis y estructura interna

Se basa en la heterogeneidad entre soportes y la homogeneidad entre individuos, y la duplicación de audiencia es independiente respecto a un tercer soporte. Se recoge íntegramente de la obra citada de Chandon⁹⁹. La probabilidades de la audiencia duplicada se podrían expresar matemáticamente como sigue:

- ✓ • La probabilidad simple de la duplicación entre dos soportes S_i y S_j , no es independiente, es decir, no es aleatoria:

$$P(S_i \cap S_j) \neq P(S_i) \cdot P(S_j) \quad \forall i \neq j$$

- ✓ • La probabilidad condicional de la duplicación entre dos soportes S_i y S_j , respecto a un tercer soporte S_k , es independiente:

$$P(S_i \cap S_j / S_k) = P(S_i / S_k) \cdot P(S_j / S_k) \quad \forall i \neq j \neq k$$

La lógica interna de este modelo se basa en que la audiencia del soporte k tiene unas características comunes, generalmente distintas del total de la población. Existe, por tanto, mayor correlación entre los individuos que constituyen la audiencia de k . Dentro de esa audiencia de k , algunos contactarán con el soporte i o con el soporte j , o con los dos a la vez: esa duplicación será independiente o, al menos, tendrá menos correlación que si nos referimos a la duplicación entre i y j referidos a la población total.

Aplicando el teorema de Bayes a la probabilidad condicional de la duplicación:

$$P(S_i \cap S_j / S_k) = \frac{P(S_i \cap S_j \cap S_k)}{P(S_k)} = \frac{P(S_i \cap S_k)}{P(S_k)} \cdot \frac{P(S_j \cap S_k)}{P(S_k)}$$

Sustituyendo en las ecuaciones anteriores y generalizando, se obtiene:

⁹⁹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 401-2.

$$P(S_i \cap S_j \cap S_k) = \frac{P(S_i \cap S_k) \cdot P(S_j \cap S_k)}{P(S_k)} = \frac{P(S_i \cap S_j) \cdot P(S_j \cap S_k)}{P(S_j)} = \frac{P(S_i \cap S_j) \cdot P(S_i \cap S_k)}{P(S_i)}$$

Se puede expresar también en términos de audiencia:

$$A_i \cap A_j \cap A_k = \frac{(A_i \cap A_k) \cdot (A_j \cap A_k)}{A_k} = \frac{(A_i \cap A_j) \cdot (A_j \cap A_k)}{A_j} = \frac{(A_i \cap A_j) \cdot (A_i \cap A_k)}{A_i}$$

En realidad no se cumplen tales identidades, ya que puede existir cierta correlación, y la triplicación se estima como la media de los tres cocientes:

$$A_i \cap A_j \cap A_k = \frac{1}{3} \cdot \left(\frac{(A_i \cap A_k) \cdot (A_j \cap A_k)}{A_k} + \frac{(A_i \cap A_j) \cdot (A_j \cap A_k)}{A_j} + \frac{(A_i \cap A_j) \cdot (A_i \cap A_k)}{A_i} \right)$$

Del mismo modo deberíamos proceder para calcular las cuádruplicaciones, las repeticiones de grado r o, finalmente, de grado n . Igual que en el modelo de agregación secuencial, el alcance neto se estima sumando las audiencias brutas (A), restando las duplicaciones (D), sumando las triplicaciones (T), restando las cuádruplicaciones (Q),..., hasta sumar (o restar) las repeticiones de orden n (N):

$$A_1 \cup A_2 \cup A_3 \cup A_4 \cup \dots \cup A_n = A - D + T - Q + \dots \pm N$$

Este modelo plantea una difícil aplicación práctica. Su cálculo se hace muy complejo, incluso con la ayuda de las hojas de cálculo. Hallar la estimación de las repeticiones de orden r , supondría resolver $\binom{n}{r} \cdot \binom{r-2}{r-2}$ multiplicaciones, sumas y divisiones. La dificultad aumenta enormemente con el número de inserciones. Por esa razón no hemos aplicado este modelo a nuestro estudio. A continuación se ejemplifica el modelo, pudiendo observarse la complejidad de cálculo, sólo con cuatro revistas.

7.2. Ejemplo

Recogemos nuevamente en la **Tabla 2.68** los datos de audiencias y duplicaciones de las cuatro revistas habituales (Muy Interesante, Teleprograma, El Jueves y Teleindiscreta) referidos al *target* de hombres, de 14 a 34 años y de clase social alta y media alta.

REVISTA	Muy Interes. (S ₁)	Teleprograma (S ₂)	El Jueves (S ₃)	Teleindiscreta (S ₄)
Muy Interes. (S ₁)	309.918	55.208	47.445	30.036
Teleprograma (S ₂)		236.902	23.935	29.574
El Jueves (S ₃)			130.265	11.555
Teleindiscreta (S ₄)				129.992

Tabla 2.68

Se deben estimar todas las posibles triplicaciones de estas revistas:

$$\begin{aligned}
 A_1 \cap A_2 \cap A_3 &= \frac{1}{3} \cdot \left(\frac{(A_1 \cap A_3) \cdot (A_2 \cap A_3)}{A_3} + \frac{(A_1 \cap A_2) \cdot (A_2 \cap A_3)}{A_2} + \frac{(A_1 \cap A_2) \cdot (A_1 \cap A_3)}{A_1} \right) = \\
 &= \frac{1}{3} \cdot \left(\frac{D_{13} \cdot D_{23}}{A_3} + \frac{D_{12} \cdot D_{23}}{A_2} + \frac{D_{12} \cdot D_{13}}{A_1} \right) = \\
 &= \frac{1}{3} \cdot \left(\frac{47.445 \cdot 23.935}{130.265} + \frac{55.208 \cdot 23.935}{236.902} + \frac{55.208 \cdot 47.445}{309.918} \right) \approx 7.582
 \end{aligned}$$

$$\begin{aligned}
 A_1 \cap A_2 \cap A_4 &= \frac{1}{3} \cdot \left(\frac{D_{14} \cdot D_{24}}{A_4} + \frac{D_{12} \cdot D_{24}}{A_2} + \frac{D_{12} \cdot D_{14}}{A_1} \right) = \\
 &= \frac{1}{3} \cdot \left(\frac{30.036 \cdot 29.574}{129.992} + \frac{55.208 \cdot 29.574}{236.902} + \frac{55.208 \cdot 30.036}{309.918} \right) \approx 6.359
 \end{aligned}$$

$$\begin{aligned}
 A_1 \cap A_3 \cap A_4 &= \frac{1}{3} \cdot \left(\frac{D_{14} \cdot D_{34}}{A_4} + \frac{D_{13} \cdot D_{34}}{A_3} + \frac{D_{13} \cdot D_{14}}{A_1} \right) = \\
 &= \frac{1}{3} \cdot \left(\frac{30.036 \cdot 11.555}{129.992} + \frac{47.445 \cdot 11.555}{130.265} + \frac{47.445 \cdot 30.036}{309.918} \right) \approx 3.825
 \end{aligned}$$

$$\begin{aligned}
 A_2 \cap A_3 \cap A_4 &= \frac{1}{3} \cdot \left(\frac{D_{24} \cdot D_{34}}{A_4} + \frac{D_{23} \cdot D_{34}}{A_3} + \frac{D_{23} \cdot D_{24}}{A_2} \right) = \\
 &= \frac{1}{3} \cdot \left(\frac{29.574 \cdot 11.555}{129.992} + \frac{23.935 \cdot 11.555}{130.265} + \frac{23.935 \cdot 29.574}{236.902} \right) \approx 2.580
 \end{aligned}$$

La cuadruplicación se estima igual con las triplicaciones:

$$\begin{aligned}
 A_1 \cap A_2 \cap A_3 \cap A_4 &= \frac{1}{6} \cdot \left(\frac{(A_1 \cap A_2 \cap A_3) \cdot (A_1 \cap A_2 \cap A_4)}{A_1 \cap A_2} + \frac{(A_1 \cap A_2 \cap A_3) \cdot (A_1 \cap A_3 \cap A_4)}{A_1 \cap A_3} + \right. \\
 &+ \frac{(A_1 \cap A_2 \cap A_4) \cdot (A_1 \cap A_3 \cap A_4)}{A_1 \cap A_4} + \frac{(A_1 \cap A_2 \cap A_3) \cdot (A_2 \cap A_3 \cap A_4)}{A_2 \cap A_3} + \frac{(A_1 \cap A_2 \cap A_4) \cdot (A_2 \cap A_3 \cap A_4)}{A_2 \cap A_4} + \\
 &+ \left. \frac{(A_1 \cap A_3 \cap A_4) \cdot (A_2 \cap A_3 \cap A_4)}{A_3 \cap A_4} \right) = \frac{1}{6} \cdot \left(\frac{7.582 \cdot 6.359}{55.208} + \frac{7.582 \cdot 3.825}{47.445} + \frac{6.359 \cdot 3.825}{30.036} + \right. \\
 &+ \left. \frac{7.582 \cdot 2.580}{23.935} + \frac{6.359 \cdot 2.580}{29.574} + \frac{3.825 \cdot 2.580}{11.555} \right) \approx 753
 \end{aligned}$$

Finalmente se estima el alcance neto de las cuatro revistas:

$$\begin{aligned}
 A_1 \cup A_2 \cup A_3 \cup A_4 &= \sum_{i=1}^4 A_i - \sum_{i=1}^3 \sum_{j=i+1}^4 (A_i \cap A_j) + \sum_{i=1}^2 \sum_{j=i+1}^3 \sum_{k=j+1}^4 (A_i \cap A_j \cap A_k) - A_1 \cap A_2 \cap A_3 \cap A_4 = \\
 &= (309.918 + 236.902 + 130.265 + 129.992) - (55.208 + 47.445 + 30.036 + 29.935 + \\
 &+ 29.574 + 11.555) + (7.582 + 6.359 + 3.825 + 2.580) - 753 \approx 628.917
 \end{aligned}$$

8. Modelo de Hofmans ampliado

8.1. Extensión a la distribución de contactos

Este modelo estocástico asume las hipótesis de homogeneidad de los individuos, heterogeneidad de los soportes y duplicación no libre. Según el modelo de Hofmans, el alcance neto de n soportes se ajusta a una serie geométrica infinita, cuya base es la audiencia bruta A , y la razón es $k \cdot D / A^{100}$:

$$C = \frac{A}{1 + k \cdot \frac{D}{A}} = A - k \cdot D + \frac{(k \cdot D)^2}{A} - \frac{(k \cdot D)^3}{A^2} + \dots \pm \frac{(k \cdot D)^{n-1}}{A^{n-2}} \pm \dots$$

$$\text{siendo } k \cdot D = \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij} \cdot D_{ij} \quad \text{y} \quad k_{ij} = \frac{A_i + A_j}{A_i + A_j - D_{ij}}$$

Si queremos cortar esa serie infinita y considerar sólo los primeros n términos, entonces la suma de las repeticiones de orden r (S_r) debería ser aproximadamente igual a:

$$S_r = \frac{(k \cdot D)^{r-1}}{A^{r-2}}$$

Esto debería ser válido si y sólo si $k \cdot D < A$, porque cuando $k \cdot D \geq A$, la serie geométrica no converge. Se puede aplicar una regla general: cuando $A_{ij} \geq (A_i + A_j)/2$ se cumplirá que $k \cdot D < A$. De todas formas, aunque $k \cdot D$ sea menor que A , el corte de la serie introduce importantes errores en el alcance neto estimado, pero éstos serán menores o iguales que $\pm(k \cdot D)^{n-1} / A^{n-2}$.

¹⁰⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 406-8.

Definiendo $\sum R_{r/n}$ como la suma del alcance neto de todas la combinaciones de r soportes entre los n . Ahora los expresamos en función de los correspondientes S_i :

$$\sum R_{1/n} = S_1$$

$$\sum R_{2/n} = \binom{n-1}{1} \cdot S_1 - S_2$$

$$\sum R_{3/n} = \binom{n-1}{2} \cdot S_1 - \binom{n-2}{1} \cdot S_2 + S_3$$

...

$$\sum R_{r/n} = \binom{n-1}{r-1} \cdot S_1 - \binom{n-2}{r-2} \cdot S_2 + \binom{n-3}{r-3} \cdot S_3 - \dots + S_r$$

Expresándolo de forma más sencilla:

$$\sum R_{r/n} = \sum_{i=1}^r (-1)^{i+1} \cdot \binom{n-i}{r-i} \cdot S_i$$

Esta ecuación permite expresar S_r en función de $\sum R_{r/n}$, que puede ser estimado por el modelo de Hofmans. S_r vendrá determinado así:

$$S_r = \sum_{i=1}^{r-1} (-1)^{i+1} \cdot \binom{n-i}{r-i} \cdot S_i - \sum R_{r/n} \quad (\text{si } r \text{ es par})$$

$$S_r = \sum_{i=1}^{r-1} (-1)^i \cdot \binom{n-i}{r-i} \cdot S_i + \sum R_{r/n} \quad (\text{si } r \text{ es impar})$$

Este método estima el mismo alcance neto que la fórmula de Hofmans, pero esta ampliación además estima la distribución de contactos: a partir del teorema de Waring, se obtiene el número de individuos expuestos a k de los n soportes:

$$E_k^n = \sum_{i=k}^n (-1)^{i-k} \binom{i}{k} \cdot S_i$$

8.2. Ejemplo

Vamos a estimar la distribución de contactos de las Revistas Muy Interesante, Teleprograma, El Jueves y Teleindiscreta. Sabíamos que la suma de sus audiencias (A) era de 807.077 personas, que coincide con la suma del alcance de las cuatro revistas tomadas de una en una ($\sum R_{1/4}$). La suma de las duplicaciones (D) valía 197.753. Calculamos ahora la suma de los alcances netos de todas las parejas, según los datos de la posterior **Tabla 2.69**:

$$\sum R_{2/4} = 491.612 + 392.738 + 409.874 + 343.232 + 337.320 + 248.702 = 2.223.478$$

Los alcances de los posibles tríos de soportes se estiman según la fórmula de Hofmans:

- Muy Interesante, Teleprograma y El Jueves:

$$AN = \frac{(309.918 + 236.902 + 130.265)^2}{677.085 + 1,1123 \cdot 55.208 + 1,1208 \cdot 47.445 + 1,0697 \cdot 23.935} \approx 560.944$$

- Muy Interesante, Teleprograma y Teleindiscreta:

$$AN = \frac{(309.918 + 236.902 + 129.992)^2}{676.812 + 1,1123 \cdot 55.208 + 1,0733 \cdot 30.036 + 1,0877 \cdot 29.574} \approx 570.720$$

- Muy Interesante, El Jueves y Teleindiscreta:

$$AN = \frac{(309.918 + 130.265 + 129.992)^2}{570.175 + 1,1208 \cdot 47.445 + 1,0733 \cdot 30.036 + 1,0465 \cdot 11.555} \approx 486.908$$

- Teleprograma, El Jueves y Teleindiscreta:

$$AN = \frac{(236.902 + 130.265 + 129.992)^2}{497.159 + 1,0697 \cdot 23.935 + 1,0877 \cdot 29.574 + 1,0465 \cdot 11.555} \approx 435.904$$

Y la suma de los alcances netos de las combinaciones de tres revistas:

$$\sum R_{3/4} = 560.944 + 570.720 + 486.908 + 435.904 = 2.054.476$$

Solamente cabe una combinación de cuatro revistas. Su alcance neto ya se calculó en el ejemplo del modelo de Hofmans (§A.4.2). Recordemos que valía 636.185, luego $\sum R_{4/4} = 636.185$.

Ya podemos hallar las sumas de las r repeticiones (S_r):

$$S_1 \text{ (suma de audiencias) } = A = 807.077$$

$$\begin{aligned} S_2 \text{ (suma de duplicaciones) } &= D = 197.753 = \\ &= \binom{3}{1} \cdot S_1 - \sum R_{2/4} = 3 \cdot 807.077 - 2.223.478 \end{aligned}$$

$$\begin{aligned} S_3 \text{ (suma de triplicaciones) } &= -\binom{3}{2} \cdot S_1 + \binom{2}{1} \cdot S_2 + \sum R_{3/4} = \\ &= -3 \cdot 807.077 + 2 \cdot 197.753 + 2.054.476 = 28.751 \end{aligned}$$

$$\begin{aligned} S_4 \text{ (suma de cuadruplicaciones) } &= \binom{3}{3} \cdot S_1 - \binom{2}{2} \cdot S_2 + \binom{1}{1} \cdot S_3 - \sum R_{4/4} = \\ &= 807.077 - 197.753 + 28.751 - 636.185 = 1.890 \end{aligned}$$

El alcance neto se mantiene igual que antes:

$$A_{1,2,3,4} = S_1 - S_2 + S_3 - S_4 = 807.077 - 197.753 + 28.751 - 1.890 = 636.185$$

Y la distribución de contactos se calcula así:

$$E_0^4 = \sum_{i=0}^4 (-1)^i \binom{i}{0} \cdot S_i = P - S_1 + S_2 - S_3 + S_4 = 818.816$$

$$E_1^4 = \sum_{i=1}^4 (-1)^{i-1} \binom{i}{1} \cdot S_i = S_1 - 2S_2 + 3S_3 - 4S_4 = 490.264$$

$$E_2^4 = \sum_{i=2}^4 (-1)^{i-2} \binom{i}{2} \cdot S_i = S_2 - 3S_3 + 6S_4 = 122.840$$

$$E_3^4 = \sum_{i=3}^4 (-1)^{i-3} \binom{i}{3} \cdot S_i = S_3 - 4S_4 = 21.191$$

$$E_4^4 = \sum_{i=4}^4 (-1)^{i-4} \binom{i}{4} \cdot S_i = S_4 = 1.890$$

9. Modelo de Metheringham

9.1. Fundamentos teóricos y desarrollo

Metheringham¹⁰¹ propuso un modelo para estimar la audiencia neta y la audiencia neta acumulada, pero ahora nos referiremos sólo a la primera. El investigador inglés aceptaba que la fórmula de Agostini proporcionaba estimaciones bastante cercanas a la realidad en las revistas americanas y francesas, pero no en las publicaciones británicas. En esos casos encontró errores de estimación que alcanzaban el 17%. Plantea una fórmula alternativa, que requiere únicamente la audiencias de las publicaciones y sus duplicaciones.

Notaba p_i como la cobertura o proporción de la población objetivo expuesta a la publicación i . Igualmente, $q_i = 1 - p_i$, es el porcentaje no expuesto a la publicación i . La no cobertura de dos soportes se denomina q_{ij} . La suma de las no coberturas de todos los pares será Σq_{ij} . Del mismo modo, Σq_{ijk} será la suma de las no coberturas de todas las combinaciones de tres soportes.

Suponiendo n publicaciones, existirán $\binom{n}{2}$ pares distintos, $\binom{n}{3}$ tríos distintos,... hasta $\binom{n}{r}$ combinaciones distintas, tomadas de r en r . Siendo k_r la media de las no coberturas de las i publicaciones:

$$k_1 = \frac{\Sigma q_i}{\binom{n}{1}} = \frac{\Sigma q_i}{n}$$

$$k_2 = \frac{\Sigma q_{ij}}{\binom{n}{2}}$$

¹⁰¹ METHERINHAM, Richard A. "Measuring the Net Cumulative Coverage of a Print Campaign". *Journal of Advertising Research*. Diciembre 1964. Vol. 4. Págs. 23-28.

$$k_3 = \frac{\sum q_{ijk}}{\binom{n}{3}}$$

...

$$k_r = \frac{\sum q_{ijk} \dots (r \text{ soportes})}{\binom{n}{r}}$$

...

$$k_n = \frac{\sum q_{ijk} \dots (n \text{ soportes})}{\binom{n}{n}} = \sum q_{ijk} \dots (n \text{ soportes})$$

El alcance neto de n soportes (C_n), se determina como el complementario a la no cobertura de los n soportes:

$$C_n = 1 - k_n = 1 - \sum q_{ijk} \dots (n \text{ soportes})$$

Los valores k_1 y k_2 se calculan a partir de las audiencias y duplicaciones de los soportes. Conociendo la secuencia restante de k , el problema estaría resuelto. Como se muestra posteriormente, el término general k_r puede aproximarse con la siguiente fórmula:

$$k_r = \frac{s \cdot (s+1) \cdot (s+2) \cdot \dots \cdot (s+r-1)}{t \cdot (t+1) \cdot (t+2) \cdot \dots \cdot (t+r-1)}$$

donde s y t toman valores tales que ningún término de la secuencia sea negativo o nulo. Esos valores se calculan a partir de k_1 , la probabilidad de no ser audiencia de un soporte medio, y k_2 , la probabilidad de no ser audiencia de una combinación media de dos soportes:

$$k_1 = \frac{s}{t} \qquad k_2 = \frac{s \cdot (s+1)}{t \cdot (t+1)}$$

Resolviendo el sistema de dos ecuaciones con dos incógnitas, se obtiene:

$$s = k_1 \cdot \frac{k_1 - k_2}{k_2 - k_1^2} \quad t = \frac{s}{k_1} = \frac{k_1 - k_2}{k_2 - k_1^2}$$

Si $k_2 = k_1^2$, los valores de s y t tienden a infinito. En este caso la duplicación es normal, y k_n es simplemente k_1^n . Probando este modelo para periódicos ingleses, descubrió que mejoraba las estimaciones de Agostini. Al combinar nueve diarios británicos en grupos de cuatro, cinco y ocho soportes, observó que la precisión aumentaba con el n° de soportes. Por su parte, Chandon confirmó esta conclusión, señalando que el modelo de Metherringham es el más preciso ante muchas publicaciones.

Metherringham¹⁰² propuso estimar la **distribución de contactos** a partir del teorema de Waring, enunciado en 1792. Dicho teorema calcula el número de contactos a partir de la suma de las probabilidades individuales (S_1), la suma de todas las probabilidades de duplicación (S_2), ..., la suma de todas las probabilidades de r -plicación (S_r), ..., hasta la suma de la probabilidad de n -plicación (S_n):

$$S_1 = \sum_{i=1}^n p_i \quad S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_{ij} \quad \dots \quad S_r = \sum_{i=1}^{n-r+1} \sum_{j=i+1}^{n-r+2} \dots \sum_{r=n-1}^n p_{ij \dots r} \quad \dots$$

$$S_n = \sum_{i=1}^1 \sum_{j=i+1}^2 \dots \sum_{n=n-1}^n p_{ij \dots n} = p_{ij \dots n}$$

El teorema establece que la probabilidad de que un individuo esté expuesto exactamente a k de las n revistas vale:

$$p(k/n; p_1, p_2, \dots, p_n) = S_k - \binom{k+1}{k} \cdot S_{k+1} + \binom{k+2}{k} \cdot S_{k+2} \dots \pm \binom{n}{k} \cdot S_n$$

cumpléndose que $0 \leq k \leq n$ y $S_0 = 1$. El número estimado de individuos que reciben exactamente k contactos, es el producto de esa probabilidad por el público objetivo:

¹⁰² METHERRINGHAM, Richard A. "Measuring the Net Cumulative Coverage of a Print Campaign". *Journal of Advertising Research*. Diciembre 1964. Vol. 4. Págs. 27-28.

$$E_k^n = P \cdot [p(k/n; p_1, p_2, \dots, p_n)] = P \cdot \left[\sum_{i=k}^n (-1)^{i-k} \cdot \binom{i}{k} \cdot S_i \right]$$

Por otra parte, la probabilidad de que un individuo esté expuesto al menos a k de los n soportes, vale:

$$p(x \geq k/n; p_1, p_2, \dots, p_n) = S_k - \binom{k}{k-1} \cdot S_{k+1} + \binom{k+1}{k-1} \cdot S_{k+2} \dots \pm \binom{n-1}{k-1} \cdot S_n$$

Luego el número esperado de individuos, expuestos al menos a k soportes, es:

$$E_{x \geq k}^n = P \cdot [p(x \geq k/n; p_1, p_2, \dots, p_n)] = P \cdot \left[\sum_{i=k}^n (-1)^{i-k} \cdot \binom{i-1}{k-1} \cdot S_i \right]$$

Como caso particular, el alcance neto C_n no es más que el número de personas que contactan al menos una vez:

$$\begin{aligned} E_{x \geq 1}^n = C_n &= P \cdot [p(x \geq 1/n; p_1, p_2, \dots, p_n)] = P \cdot [1 - p(0/n; p_1, p_2, \dots, p_n)] = \\ &= P \cdot [S_1 - S_2 + S_3 - \dots \pm S_n] \end{aligned}$$

Como señala Chandon¹⁰³, las probabilidades individuales de exposición, por hipótesis, siguen una distribución Beta y, por tanto, la distribución de contactos, resultante de n inserciones en un hipotético soporte medio, es Beta-Binomial, porque la exposición se considera una variable Bernoulli. Por tanto, ahora el problema es estimar los parámetros a y b de la distribución Beta, en función de las audiencias individuales y las duplicaciones. Sean, pues, \bar{p} y \bar{D} , respectivamente, las probabilidades medias de las audiencias individuales y de las duplicaciones:

¹⁰³ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 418-22.

$$\bar{p} = \frac{\sum_{i=1}^n p_i}{n} \quad \bar{D} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n p_{ij}}{\binom{n}{2}}$$

Como la probabilidad media individual \bar{p} no es más que la probabilidad de un hipotético soporte medio, la distribución Beta Binomial nos da la probabilidad de exposición exactamente a una inserción:

$$\bar{p} = p(1/1; \bar{p}) = \frac{a}{a+b}$$

De igual modo, la probabilidad media de duplicación \bar{D} sería la probabilidad de contactar dos veces en dos inserciones, en un hipotético soporte medio:

$$\bar{D} = p(2/2; \bar{D}) = \frac{(a+1) \cdot a}{(a+b+1) \cdot (a+b)}$$

Resolviendo las dos ecuaciones anteriores, obtenemos los valores estimados de los parámetros a y b :

$$a = \frac{\bar{p}^2 - \bar{p} \cdot \bar{D}}{\bar{D} - \bar{p}^2} \quad b = \frac{a \cdot (1 - \bar{p})}{\bar{p}}$$

Determinados los parámetros a y b , la probabilidad de recibir k contactos entre n inserciones será igual a:

$$P(k, a, b, n) = \binom{n}{k} \cdot \frac{(a+k-1) \cdot (a+k-2) \cdot \dots \cdot a \cdot (b+n-k-1) \cdot (b+n-k-2) \cdot \dots \cdot b}{(a+b+n-1) \cdot (a+b+n-2) \cdot \dots \cdot (a+b)}$$

Y el número de individuos expuestos a k de las n inserciones será:

$$E_k^n = P \cdot p(k, a, b, n)$$

Chandon demostró que los parámetros a y b pueden ser estimados directamente a partir de la suma de audiencias y de la suma de las duplicaciones, sin recurrir ni a la

audiencia ni a la duplicación media. Definiendo la media y la varianza de una distribución Beta-Binomial, y despejando a y b :

$$\begin{cases} \mu = \frac{n \cdot a}{a + b} \\ \sigma^2 = \frac{n \cdot a \cdot b \cdot (n + a + b)}{(a + b)^2 \cdot (a + b + 1)} \end{cases} \Rightarrow \begin{cases} a = \frac{\mu \cdot [\sigma^2 - (n - \mu)]}{\mu \cdot (n - \mu) - n \cdot \sigma^2} \\ b = a \cdot \frac{n - \mu}{\mu} \end{cases}$$

Expresando la media μ y la varianza σ^2 en función de las audiencias y las duplicaciones:

$$\begin{cases} \mu = \frac{A}{P} & \text{siendo } A = \sum_{i=1}^n A_i \\ \sigma^2 = \frac{A + 2D}{P} - \left(\frac{A}{P} \right)^2 & \text{siendo } D = \sum_{i=1}^{n-1} \sum_{j=i+1}^n A_{ij} \end{cases}$$

Remplazando estos valores de la media y la varianza en las ecuaciones anteriores, y después de varias operaciones, obtenemos la estimación de los parámetros a y b :

$$\begin{cases} a = \frac{2DA - A^2(n-1)}{A^2(n-1) - 2DnP} \\ b = a \frac{nP - A}{A} \end{cases}$$

Los parámetros a y b toman los mismos valores, independientemente de su forma de estimación. Chandon muestra con ello que el modelo de Metheringham se reduce a la estimación de a y b por el método de los momentos, bajo la hipótesis de que las exposiciones siguen una distribución Beta-Binomial.

9.2. Ejemplo

Empezaremos hallando el alcance neto de las parejas de revistas de nuestro ejemplo habitual, tal como recoge la **Tabla 2.69**:

Revista (i)	A _i	Revista (j)	A _j	D _{ij}	AN = A _i + A _j - D _{ij}
Muy Interes.	309.918	Teleprograma	236.902	55.208	491.612
Muy Interes.	309.918	El Jueves	130.265	47.445	392.738
Muy Interes.	309.918	Teleindiscreta	129.992	30.036	409.874
Teleprograma	236.902	El Jueves	130.265	23.935	343.232
Teleprograma	236.902	Teleindiscreta	129.992	29.574	337.320
El Jueves	130.265	Teleindiscreta	129.992	11.555	248.702

Tabla 2.69

Hallamos primeramente las probabilidades de no contactar con uno de los soportes: $q_i = 1 - p_i$.

$$q_1 = 1 - p_1 = 1 - \frac{309.918}{1.455.001} \approx 0,787 \quad q_2 = 1 - p_2 = 1 - \frac{236.902}{1.455.001} \approx 0,8372$$

$$q_3 = 1 - p_3 = 1 - \frac{130.265}{1.455.001} \approx 0,9105 \quad q_4 = 1 - p_4 = 1 - \frac{129.992}{1.455.001} \approx 0,9107$$

Luego k_1 toma el siguiente valor:

$$k_1 = \frac{\sum_{i=1}^n q_i}{n} = \frac{0,787 + 0,8372 + 0,9105 + 0,9107}{4} \approx 0,8613$$

Hallemos ahora la no cobertura de los pares de soportes, es decir la proporción de los que no leen ninguno de los dos ($q_{ij} = 1 - p_{ij}$):

$$q_{12} = 1 - p_{12} = 1 - \frac{491.612}{1.455.001} \approx 0,6621 \quad q_{13} = 1 - p_{13} = 1 - \frac{392.738}{1.455.001} \approx 0,7301$$

$$q_{14} = 1 - p_{14} = 1 - \frac{409.874}{1.455.001} \approx 0,7183 \quad q_{23} = 1 - p_{23} = 1 - \frac{343.232}{1.455.001} \approx 0,7641$$

$$q_{24} = 1 - p_{24} = 1 - \frac{337.320}{1.455.001} \approx 0,7682 \quad q_{34} = 1 - p_{34} = 1 - \frac{248.702}{1.455.001} \approx 0,8291$$

Determinamos la media de las no-coberturas de las parejas de soportes:

$$k_2 = \frac{\sum q_{ij}}{\binom{4}{2}} = \frac{0,6621 + 0,7301 + 0,7183 + 0,7641 + 0,7682 + 0,8291}{\frac{4 \cdot (4-1)}{2}} \approx 0,7452$$

A partir de k_1 y k_2 , podemos calcular s y t :

$$t = \frac{k_1 - k_2}{k_2 - k_1^2} = \frac{0,8613 - 0,7452}{0,7452 - (0,8613)^2} \approx 34$$

$$s = k_1 \cdot t = 0,8613 \cdot 34 \approx 29,3$$

Finalmente se determina el porcentaje de los que no contactan con ninguno de los cuatro soportes (k_4) y, a través de ello, el alcance neto:

$$k_4 = \frac{s \cdot (s+1) \cdot (s+2) \cdot \dots \cdot (s+3)}{t \cdot (t+1) \cdot (t+2) \cdot \dots \cdot (t+3)} = \frac{29,3 \cdot 30,3 \cdot 31,3 \cdot 32,3}{34 \cdot 35 \cdot 36 \cdot 37} \approx 0,565$$

$$C_4 = (1 - k_4) \cdot P = (1 - 0,565) \cdot 1.455.001 \approx 632.856$$

Se calcula ahora la **distribución de contactos**, siguiendo el proceso expuesto por Chandon:

$$\bar{p} = \frac{\sum_{i=1}^4 p_i}{4} = \frac{\sum_{i=1}^4 A_i}{4 \cdot P} = \frac{807.077}{4 \cdot 1.455.001} \approx 0,139$$

$$\bar{D} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n p_{ij}}{\binom{4}{2}} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n D_{ij}}{\binom{4}{2} \cdot P} = \frac{197.753}{6 \cdot 1.455.001} \approx 0,023$$

$$a = \frac{\bar{p}^2 - \bar{p} \cdot \bar{D}}{\bar{D} - \bar{p}^2} = \frac{0,139^2 - 0,139 \cdot 0,023}{0,023 - 0,139^2} \approx 4,7$$

$$b = \frac{a \cdot (1 - \bar{p})}{\bar{p}} = \frac{4,7 \cdot (1 - 0,139)}{0,139} \approx 29,2$$

Determinados los parámetros, la distribución de contactos se estima así:

$$E_0^4 = \binom{4}{0} \cdot \frac{32,1 \cdot 31,1 \cdot 30,1 \cdot 29,1}{36,8 \cdot 35,8 \cdot 34,8 \cdot 33,8} \cdot 1.455.001 \approx 0,5643 \cdot 1.455.001 \approx 821.057$$

$$E_1^4 = \binom{4}{1} \cdot \frac{4,7 \cdot 31,1 \cdot 30,1 \cdot 29,1}{36,8 \cdot 35,8 \cdot 34,8 \cdot 33,8} \cdot 1.455.001 \approx 4 \cdot 0,083 \cdot 1.455.001 \approx 480.854$$

$$E_2^4 = \binom{4}{2} \cdot \frac{5,7 \cdot 4,7 \cdot 30,1 \cdot 29,1}{36,8 \cdot 35,8 \cdot 34,8 \cdot 33,8} \cdot 1.455.001 \approx 6 \cdot 0,015 \cdot 1.455.001 \approx 132.181$$

$$E_3^4 = \binom{4}{3} \cdot \frac{6,7 \cdot 5,7 \cdot 4,7 \cdot 29,1}{36,8 \cdot 35,8 \cdot 34,8 \cdot 33,8} \cdot 1.455.001 \approx 4 \cdot 0,003 \cdot 1.455.001 \approx 19.612$$

$$E_4^4 = \binom{4}{4} \cdot \frac{7,7 \cdot 6,7 \cdot 5,7 \cdot 4,7}{36,8 \cdot 35,8 \cdot 34,8 \cdot 33,8} \cdot 1.455.001 \approx 0,001 \cdot 1.455.001 \approx 1.297$$

La suma de los que reciben 0, 1, 2, 3 y 4 contactos, debe ser el total de la población:

$$P = E_0^4 + E_1^4 + E_2^4 + E_3^4 + E_4^4 = 821.057 + 480.854 + 132.181 + 19.612 + 1.297 = 1.455.001$$

9.3. Tests del alcance neto estimado

El de Metherringham es el modelo estocástico que mejor ha estimado el alcance neto. Tiende a subestimar ligeramente el alcance, como muestra que la media de errores sea -1,57%; los errores, en uno o en otro sentido, alcanzan como media el 1,69% del alcance neto. Mejora incluso el modelo de Hofmans cuya media absoluta de errores era de 1,84%, y se asemeja al de Agostini (1,64%) y de Kaatz (1,65%). Hemos encontrado variaciones entre los distintos públicos objetivo, pero sin que se pueda establecer una relación con el sexo, la edad o la clase social (cfr. **Tabla 2.48**). Volvemos a encontrar un segmento con errores superiores a la media: los hombres, mayores de 54 años, de clase social media media. Lo mismo sucedía en los modelos de Agostini, Hofmans y Kaatz,

precisamente los que mejor estiman. No se trata, por tanto, de plantearse errores de cálculo o de recogida de datos, ni de buscar una explicación aleatoria, precisamente porque hay errores sistemáticos en ese *target*. Como se puede ver en la **Tabla 2.3**, el segmento tiene unos tamaños de muestra y de población representativos (otros menores no dan problemas). Debemos concluir que ese segmento se comporta de forma atípica, sin seguir la tendencia general. Pero si hemos encontrado un grupo atípico, podría suceder lo mismo con los *targets* de futuras campañas. No obstante, el problema se ha de plantear en su verdadera dimensión: habitualmente hemos encontrado unos errores que rondan el 1'5% del alcance, y en uno de los 19 grupos, esos errores se han elevado al 4'1%, margen que quizá sea aceptado por muchos.

POBLACIÓN OBJETIVO	MEDIA e%	DESV. EST. e%	MEDIA e% 	DESV. EST. e%
H, 14-34, A-MA	-1,1627	0,7231	1,1874	0,6795
H, 14-34, MM	-1,7615	0,9382	1,7615	0,9382
H, 14-34, MB-B	-1,3713	0,9317	1,3864	0,9080
H, 35-54, A-MA	-1,3528	1,1801	1,4730	1,0174
H, 35-54, MM	-1,4334	1,0914	1,5130	0,9716
H, 35-54, MB-B	-1,5436	1,4257	1,6362	1,3125
H, >54, A-MA	-1,6588	0,8973	1,7055	0,7998
H, >54, MM	-4,1533	1,9903	4,1533	1,9903
H, >54, MB-B	-1,1651	0,8106	1,2141	0,7311
M, 14-34, A-MA	-1,3013	1,0582	1,3709	0,9611
M, 14-34, MM	-2,1642	1,2959	2,2025	1,2261
M, 14-34, MB-B	-2,3910	1,3161	2,4751	1,1406
M, 35-54, A-MA	-2,5428	1,5927	2,6272	1,4413
M, 35-54, MM	-1,9298	0,7825	1,9298	0,7825
M, 35-54, MB-B	-1,8102	1,0934	1,8223	1,0719
M, >54, A-MA	0,3911	1,1239	1,0530	0,5070
M, >54, MM	-0,5675	0,4358	0,6022	0,3836
M, >54, MB-B	-0,1779	0,5188	0,3341	0,4301
TOTAL	-1,6689	0,7353	1,6905	0,6827
TODOS LOS CASOS	-1,5722	1,3959	1,6914	1,2483

Tabla 2.70: Errores del modelo de Metheringham, según el público objetivo.

En cuanto a los criterios de selección de soportes, se observa en la **Tabla 2.71** que tienen distintos márgenes de error en valor absoluto: la cobertura es el criterio que peor estima, produciendo un error medio del 2%; le siguen la rentabilidad (1,5%) y la afinidad (1,4%). Existen pues diferencias, aunque poco relevantes, en la estimación debido a un factor extrínseco a la propia campaña.

CRITERIO	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
rentabilidad	-1,4318	1,2756	1,5537	1,1223
cobertura	-1,8989	1,4367	1,9888	1,3081
rentabilidad+cobertura	-1,6600	1,3765	1,7474	1,2626
afinidad	-1,2553	1,4340	1,4422	1,2436
TODOS LOS CASOS	-1,5722	1,3959	1,6914	1,2483

Tabla 2.71: Errores del modelo de Metheringham, según el criterio de selección de soportes.

Los errores absolutos han aumentado con el número de soportes (**Tabla 2.72**), al igual que en los modelos de Agostini, Hofmans y Kaatz. Siguen una evolución parecida a la de Hofmans (**Tabla 2.34**): con pocas inserciones la estimación es muy buena, pero con muchas se hace peor que las demás. El modelo de Metheringham es el que mejor ha estimado la audiencia neta para 4 soportes, con un error absoluto de apenas el 0,5%. Pero con 20 soportes tiene un margen del 2,5%, frente al 2% de Agostini (**Tabla 2.7**) ó al 2,2% de Kaatz (**Tabla 2.29**). La media de errores muestra que el modelo de Metheringham tiende a subestimar ligeramente todas las alternativas. La dispersión de los errores, medida a través de la desviación estándar, también se hace mayor con el número de soportes. Para 4 soportes, este modelo también es el que tiene los errores más concentrados, con una desviación del error absoluto del 0,4%.

NÚMERO DE SOPORTES	MEDIA e%	DESV. EST. e%	MEDIA e%	DESV. EST. e%
4	-0,2681	0,6090	0,5077	0,4271
8	-1,2735	0,8815	1,2870	0,8614
12	-1,8406	1,3066	1,9072	1,2059
16	-2,1058	1,5177	2,2342	1,3186
20	-2,3071	1,5988	2,4982	1,2749
TODOS LOS CASOS	-1,5722	1,3959	1,6914	1,2483

Tabla 2.72: Errores del modelo de Metheringham, según el número de soportes.

9.4. Contrastes sobre la distribución de contactos

Pretendemos aquí comparar las distribuciones de contactos estimadas por Metheringham y por el TOM2. Se testan las 375 campañas, midiendo las diferencias campaña a campaña, como antes se explicaba (§B.1.4). Se ha de notar nuevamente que estamos comparando estimaciones, sujetas en ambos casos a error (**I Parte, §D.3**). Las diferencias positivas se compensan con las negativas, como muestra la segunda columna de la **Tabla 2.73**; en valor absoluto, la diferencia general ronda el 5%, la mitad que el modelo Binomial. Metheringham y TOM estiman distribuciones bastante distintas, pero sus diferencias no siguen tendencia definible por el *target*.

POBLACIÓN OBJETIVO	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
H, 14-34, A-MA	-0,70%	0,49%	4,36%	1,43%	20
H, 14-34, MM	-1,03%	0,68%	5,13%	1,88%	20
H, 14-34, MB-B	-0,74%	0,58%	4,56%	1,96%	20
H, 35-54, A-MA	-0,65%	0,58%	4,28%	1,75%	20
H, 35-54, MM	-0,67%	0,53%	4,02%	1,92%	20
H, 35-54, MB-B	-0,50%	0,51%	2,94%	1,54%	20
H, >54, A-MA	-0,66%	0,42%	3,90%	1,44%	20
H, >54, MM	-1,67%	0,91%	6,39%	2,45%	20
H, >54, MB-B	-0,25%	0,18%	1,76%	0,83%	20
M, 14-34, A-MA	-0,81%	0,70%	5,07%	2,21%	20
M, 14-34, MM	-1,43%	0,98%	6,60%	2,91%	20
M, 14-34, MB-B	-1,58%	0,93%	7,20%	2,73%	20
M, 35-54, A-MA	-1,51%	1,06%	8,04%	3,09%	20
M, 35-54, MM	-1,14%	0,56%	6,89%	2,13%	20
M, 35-54, MB-B	-0,81%	0,54%	5,05%	1,97%	20
M, >54, A-MA	0,28%	0,60%	3,47%	1,75%	20
M, >54, MM	-0,25%	0,19%	3,67%	0,87%	20
M, >54, MB-B	-0,03%	0,11%	1,78%	0,66%	20
TOTAL	-0,73%	0,49%	2,26%	1,11%	15
TODOS LOS CASOS	-0,78%	0,81%	4,63%	2,59%	375

Tabla 2.73: Contrastes entre las distribuciones estimadas por los modelos de Metheringham y TOM (según el público objetivo).

El criterio de selección de soportes tampoco influye en las diferencias entre las distribuciones estimadas (Tabla 2.74): las medias de los grupos son muy parecidas, y las desviaciones estándar de los grupos son similares a la desviación estándar total.

CRITERIOS	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
rentabilidad	-0,71%	0,74%	4,32%	2,38%	95
cobertura	-0,97%	0,88%	5,23%	2,78%	95
rentabilidad+cobertura	-0,85%	0,83%	4,81%	2,65%	95
afinidad	-0,59%	0,74%	4,14%	2,42%	90
TODOS LOS CASOS	-0,78%	0,81%	4,63%	2,59%	375

Tabla 2.74: Contrastes entre las distribuciones estimadas por los modelos de Metheringham y TOM (según el criterio de selección de soportes).

Como ya sucedía con el modelo Binomial, el nº de soportes determina el comportamiento de las diferencias. Según se descende en la **Tabla 2.75**, Metheringham estima frecuencias menores que el TOM y progresivamente van divergiendo más. Es paradójico en comparación con anteriores resultados, porque al aumentar las *inserciones* mejoraba el TOM y empeoraba Metheringham, hasta situarse en niveles similares (ver **Tabla 1.15**); pero aquí la inserción adicional supone un soporte adicional (sólo una inserción por soporte) y la relación con los *soportes* ya antes se planteaba más difusa.

NÚMERO DE SOPORTES	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n (f_i^B - f_i^T)$	$\sum_{i=1}^n f_i^B - f_i^T $	$\sum_{i=1}^n f_i^B - f_i^T $	Nº DE CASOS
	MEDIA	DESV. EST.	MEDIA	DESV. EST.	
4	-0,08%	0,19%	2,26%	1,55%	75
8	-0,56%	0,44%	4,28%	1,91%	75
12	-0,89%	0,70%	5,03%	2,34%	75
16	-1,13%	0,89%	5,57%	2,66%	75
20	-1,27%	0,93%	6,01%	2,54%	75
TODOS LOS CASOS	-0,78%	0,81%	4,63%	2,59%	375

Tabla 2.75: Contrastes entre las distribuciones estimadas por los modelos de Metheringham y TOM (según el nº de soportes).

C) ANÁLISIS COMPARATIVO Y CONCLUSIONES

En este apartado se comparan los modelos que estiman el alcance neto. No incluye los modelos de duplicaciones (Little-Lodish, Goodhardt-Ehrenberg, etc) ni los de excesiva complejidad de cálculo (Agregación Secuencial e Independencia Condicional). En total se examinan diez modelos, la mitad deterministas y la mitad estocásticos. De la fórmula de Agostini se recogen las cinco variantes anteriormente referidas.

El análisis se basa en varias medidas estadísticas de los errores de estimación. El error sigue computándose como la diferencia entre el alcance estimado y el observado en el TOM2, todo ello dividido por el alcance observado en el TOM2. Téngase en cuenta que esta función del error está acotada inferiormente (-100%), mientras que su límite superior es infinito. Las medidas estadísticas indican el comportamiento de los errores: la media aritmética muestra la tendencia central; la desviación estándar, la dispersión; el rango, el recorrido; el mínimo, el error más pequeño; el máximo, el mayor error; y la suma de errores, el volumen total de errores. Obviamente la media aritmética multiplicada por el número de casos (402 campañas) será igual a la suma de errores; y el rango es la diferencia entre el máximo y el mínimo.

Los resultados se resumen en las **Tablas 2.76 y 2.77**. La diferencia radica en que la segunda tabla recoge los *valores absolutos* del porcentaje de errores, para eliminar el efecto del signo.

MODELO	MEDIA e%	DESV. EST. e%	RANGO e%	MÍN. e%	MÁX. e%	SUMA e%
Agostini ($k=1,125$)	-2,92	2,18	10,30	-9,67	0,63	-1.175
Agostini ($k=1,048$)	-0,24	2,04	10,75	-6,88	3,87	-96
Agostini ($k=1,044$)	-0,04	2,04	10,80	-6,67	4,13	-17
Agostini (k var. p.o.)	-0,22	1,05	5,78	-3,64	2,14	-90
Agostini (k var. n° sop.)	0,14	1,96	11,18	-6,01	5,18	58
Kuhn	13,62	8,22	29,34	0,05	29,39	5.474
Kaatz	-0,20	2,04	10,71	-6,85	3,86	-79
Starch	-5,93	3,44	14,75	-14,25	0,50	-2.383
Hofmans	-1,44	1,91	12,26	-7,34	4,93	-580
Binomial	15,45	6,61	63,99	0,63	64,62	6.209
Sainsbury	16,27	6,19	31,72	0,90	32,61	6.541
Sainsbury modificado	10,04	6,03	71,75	-3,25	68,50	4.038
Kwerel	-7,04	7,28	25,37	-20,52	4,85	-2.831
Metheringham	-1,57	1,40	7,96	-6,14	1,82	-632

Tabla 2.76: Comportamiento del error en los modelos de estimación del alcance neto.

Resulta patente la primacía de los modelos ad-hoc, salvo en las sobrestimaciones del modelo de Kuhn. De los modelos estocásticos solamente es aceptable el de Metheringham, el que tenía hipótesis menos restrictivas. Parece que en cuanto se admite la homogeneidad de los individuos, los modelos empeoran sus estimaciones. Sin embargo, a mi juicio no se puede establecer una verdadera relación causal entre las hipótesis implícitas de un modelo y los resultados que produce: existen muchos otros factores que influyen en el comportamiento de los errores.

No se puede negar tampoco que al aceptar hipótesis más realistas, la estimación mejora. El modelo binomial, con una sobrestimación del 16%, pasa a ser del 15% con el de Sainsbury al aceptar la heterogeneidad de los soportes. A su vez, cuando se acepta la interdependencia de las audiencias, el modelo de Sainsbury modificado reduce las sobrestimaciones hasta un 10%. Sin embargo, la desviación típica es muy parecida para los dos modelos de Sainsbury. Pero en cuanto al rango, Sainsbury tiene un 68% y Sainsbury modificado un 33%, porque el segundo tiene un error máximo mucho mayor. Ello puede ser una muestra de que, al cumplirse hipótesis más realistas, no se producen resultados más satisfactorios.

El modelo de Kwerel no admite la interdependencia de las audiencias, pero es el único estocástico con Metheringham, que acepta la heterogeneidad de los individuos. Produce una subestimación media del 7%, margen de error que sólo es reducido por Metheringham. De esta manera, parece que la heterogeneidad de los individuos es la hipótesis más influyente para la precisión de los modelos estocásticos.

MODELO	MEDIA e%	DESV. EST. e%	RANGO e%	MÍN. e%	MÁX. e%	SUMA e%
Agostini ($k=1,125$)	2,94	2,15	9,65	0,02	9,67	1.183
Agostini ($k=1,048$)	1,64	1,23	6,87	0,00	6,88	661
Agostini ($k=1,044$)	1,66	1,19	6,66	0,01	6,67	667
Agostini (k var. p.o.)	0,88	0,61	3,64	0,00	3,64	355
Agostini (k var. nº sop.)	1,50	1,27	6,00	0,00	6,01	602
Kuhn	13,62	8,22	29,34	0,05	29,39	5.474
Kaatz	1,65	1,22	6,85	0,00	6,85	664
Starch	5,93	3,44	14,24	0,00	14,25	2.385
Hofmans	1,85	1,53	7,34	0,00	7,34	743
Binomial	15,45	6,61	63,99	0,63	64,62	6.209
Sainsbury	16,27	6,19	31,72	0,90	32,61	6.541
Sainsbury modificado	10,06	6,00	68,29	0,21	68,50	4.045
Kwerel	8,35	5,73	20,34	0,18	20,52	3.356
Metheringham	1,69	1,25	6,13	0,00	6,14	680

Tabla 2.77: Comportamiento del error absoluto en los modelos de alcance neto.

El modelo ad-hoc con menor error medio (-0'04%) es el de Agostini con $k=1'044$, pero ello se debe a que ese parámetro fue determinado como un valor medio para las 402 campañas. Por eso su error medio absoluto (1'66%), que no está afectado por el signo, es mayor que el obtenido con el parámetro $k=1'048$ (1'64%), determinado por la regresión lineal. En valor absoluto, el error medio más bajo (0'88%) es el del método de Agostini con k variable según la población objetivo. Le sigue el método de Agostini con k variable según el número de soportes (1'5%), pero hay que notar que considera sólo 5 alternativas (4, 8, 12, 16 y 20 revistas) en comparación con las 19 alternativas de públicos objetivos: eso hace que sea más preciso el ajuste. Como se esperaba la media absoluta de errores es menor cuando k se considera variable, aunque no tiene porque

ocurrir igual con la dispersión, como nos indica la desviación estándar de $k=1'044$ (1'19%) ó $k=1'048$ (1'23%) en comparación k variable según el número de soportes (1'27%).

Todas las variantes de $k=1'125$ de la fórmula de Agostini, han mejorado sustancialmente el modelo. Con una subestimación media de casi el 3%, la suma de errores casi duplicaba la de cualquiera de las variantes. Esto justifica que se haya buscado un valor del parámetro k para nuestro país.

El modelo de Agostini con k variable según el *target* tiene su peor estimación, entre los 402 casos, con una subestimación del 3'64%. Cuando k es constante ronda el 7%, que también es un valor pequeño teniendo en cuenta que es el caso más desfavorable. En contrapartida, sus mejores estimaciones cometen errores nulos.

El modelo de Kaatz, que pretendía hacer más preciso el modelo de Agostini, no ha conseguido más que equipararse a los casos en que k es constante, para el caso de España. El modelo de Starch ha originado una subestimación media casi del 6%, con una dispersión del 3'4%; su mejor estimación produce un error nulo, pero la peor provoca una subestimación del 14%, cifra que ya es relativamente considerable. Como además este método tiene un cálculo complejo, se hace desaconsejable su utilización.

El modelo de Hofmans, contra lo escrito en la literatura, no ha mejorado la fórmula de Agostini. Con una tendencia general a la subestimación (error medio igual a -1'44%) tiene un error medio absoluto mayor que Agostini, con el parámetro k constante. Su peor subestimación es del 7'34% y su peor sobrestimación es del 4'93%; entre esos extremos están todos los errores.

PARTE III

LA AUDIENCIA ACUMULADA

INTRODUCCIÓN

En esta III Parte se recogen diversos modelos de acumulación de audiencia. Su objetivo es medir el alcance acumulado y la distribución de contactos, ante varias exposiciones de un mismo soporte; la acumulación se da entre los números de una misma revista, los ejemplares de un mismo periódico, las cuñas de un programa de radio, los anuncios en una franja horaria de TV, etc. La velocidad de acumulación de audiencia, según Surmanek¹⁰⁴, es excelente en TV, Diarios y Suplementos, buena en Radio, regular en el Medio Exterior, y mala en Revistas. Dentro de cada medio existen diferencias: por ejemplo, la audiencia de una revista se acumula más rápidamente cuanto [1] menor sea la periodicidad entre número y número, [2] menor sea el número de lectores por ejemplar, [3] contenga más datos o noticias de actualidad, [4] mayor sea el porcentaje de ventas en quioscos y [5] mayor sea el porcentaje de lectura dentro del hogar. En los siguientes epígrafes me refiero a la acumulación de audiencia que producen sucesivas inserciones en un soporte genérico.

Desde un punto de vista práctico tiene menos importancia que la II Parte, ya que estos modelos requieren datos no disponibles. El EGM no estudia la evolución de la audiencia en los entrevistados, y no se conoce cómo se va acumulando la audiencia de los medios; para eso haría falta un panel o encuestas repetitivas sobre individuos fijos; lo que presenta, en cambio, son audiencias probabilizadas. Los autores de estos modelos disponían de la acumulación de dos, tres,..., números de una misma revista y, sobre ello, buscaban estimar el alcance y la distribución de contactos de cualquier combinación.

Esas razones imposibilitan un contraste empírico de los modelos, como se hizo en la II Parte. Aunque tenga menos interés, conviene estudiar la acumulación por tres razones: primero por el desarrollo histórico de las investigaciones de medios; segundo, por la estructura de este trabajo; tercero, y más importante, porque cualquier programa

¹⁰⁴ SURMANEK, Jim. "Una guía fácil y rápida para la Planificación de Medios". Ed. Eresma. Madrid. 1982. Págs. 12 y 65.

de planificación, asociado a sus fuentes de datos, tiene que optar por alguno de los modelos de acumulación. En este sentido, en la simulación se explica la alternativa empleada por los programas TOM, con los datos del EGM.

La clasificación expuesta se debe a Chandon¹⁰⁵, quien además elaboró una serie de criterios que permiten evaluar los modelos de acumulación. Son de tres tipos: los *teóricos*, que hacen referencia al número y naturaleza de las hipótesis, que han de ser explícitas, testables, coherentes y sencillas; los *empíricos* en función de la adecuación de las estimaciones con las observaciones reales; y los *prácticos* que evalúan los datos necesarios y a la facilidad de cálculo. En total son 15 criterios, formulados como preguntas:

CRITERIOS TEÓRICOS:

1. ¿Es Determinístico o Estocástico el fenómeno de la exposición? D/E
2. ¿Es estacionaria la exposición entre los números? sí/no/mix
3. ¿Es estacionaria la exposición entre los individuos? sí/no/mix
4. ¿Son homogéneos los individuos? sí/no
5. ¿Es puramente aleatoria la acumulación? sí/no
6. ¿Estima la distribución de contactos? sí/no
7. ¿Reproduce los datos que han servido para estimar sus parámetros? sí/no
8. ¿Tiene solución única? sí/no

CRITERIOS EMPÍRICOS:

9. ¿Son rigurosas las condiciones de aplicación? sí/no
10. ¿Es consistente el modelo? sí/no
11. ¿Las estimaciones del modelo tienen sesgos sistemáticos? +/- (sobrestimaciones/subestimaciones).
12. ¿El error medio de estimación es menor que el error muestral? sí/no
13. Si se responde afirmativamente al criterio 6, ¿estima bien la distribución de contactos, sin desplazar o suavizar los picos? sí/no

¹⁰⁵ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 293-300.

CRITERIOS PRÁCTICOS:14. ¿Contiene gran cantidad de cálculos? sí/no15. ¿Requiere más datos que las acumulaciones de dos números? sí/no/mix.

Estos criterios sirven para evaluar los modelos alternativos, de manera que el planificador pueda elegir aquel que mejor se adecúa a sus objetivos, necesidades y recursos; en la **Tabla 3.1** Chandon reunía todas las respuestas.

Modelos	Criterios teóricos								Criterios empíricos					Criter. práctico.	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Logaritmico	D	sí	sí	sí	no	no	sí	sí	no	no	-	sí	()	no	no
Exponencial	D	sí	sí	sí	no	no	sí	sí	no	no	-	no	()	no	sí
Agostini	D	sí	sí	sí	no	no	sí	sí	no	sí	no	sí	()	no	no
Duplicación Constante	D	sí	sí	sí	no	no	sí	sí	no	sí	-	no	()	no	no
Hofmans	D	sí	sí	sí	no	no	sí	sí	no	sí	+	no	()	no	sí
Binomial	E	sí	sí	sí	sí	sí	sí	sí	no	sí	+	no	no	no	no
2 segmentos	E	sí	sí	no	no	sí	sí	sí	no	no	-	no	no	no	no
3 segmentos	E	sí	sí	no	no	sí	sí	no	no	sí	no	sí	sí	no	no
Heterogeneidad lineal	E	sí	sí	no	no	sí	sí	sí	no	no	+	no	no	no	no
Beta Binomial	E	sí	sí	no	no	sí	sí	sí	no	sí	-	sí	sí	no	no
Binomial Negativa	E	sí	sí	no	no	sí	sí	sí	no	no	+	no	no	no	no
Broadbent 2 segment.	E	sí	sí	no	no	sí	sí	sí	no	sí	+	no	no	no	no
Morgensztern	E	sí	sí	no	no	sí	sí	sí	sí	sí	-	no	no	no	no
2 seg. Beta Binomial	E	sí	sí	no	no	sí	sí	sí	sí	sí	no	sí	sí	no	sí
3 seg. Beta Binomial	E	sí	sí	no	no	sí	sí	sí	sí	sí	no	sí	sí	no	sí
Markov Binomial	E	no	sí	sí	no	sí	sí	sí	no	sí	+	no	no	sí	no
Actuarial	E	s/n	s/n	no	no	sí	no	sí	no	no	-	no	no	sí	sí
Probabilidades fijas	E	s/n	s/n	no	no	sí	sí	sí	no	no	-	no	no	sí	sí
Empírico	E	s/n	s/n	no	no	sí	sí	no	no	sí	+/-	no	no	sí	sí
Personal	E	s/n	s/n	no	no	sí	sí	sí	no	sí	-	sí	sí	sí	s/n
Ultimo periodo	E	s/n	s/n	no	no	sí	sí	sí	sí	sí	+	sí	sí	sí	sí

Tabla 3.1: Evaluación de los Modelos de Acumulación. Fuente: Chandon (op. cit.).

Sin pretender una recopilación completa, a continuación se exponen varios de esos modelos, representativos de los distintos enfoques de análisis: modelos empíricos, estocásticos y de simulación.

A) MODELOS AD-HOC

Los Modelos Empíricos, Ad-Hoc o Determinísticos, como se dijo en la II Parte, se fundamentan en la observación directa de la realidad. Se desarrollaron en los años 50 y 60, a partir de los datos disponibles sobre la acumulación de audiencia de varios soportes; esa información se recogía mediante encuestas personales y luego se presentaba de forma agregada. A través de la regresión estadística y otras técnicas, buscaban relaciones que se ajustaran fielmente a los datos, para extenderlas a otros objetivos, como la medición del alcance acumulado de un soporte durante cualquier periodo. A diferencia de otros enfoques, estos modelos no calculan la distribución de contactos.

Cinco modelos se recogen en las siguientes páginas: el *logarítmico* y el *exponencial*, que son una muestra del enfoque determinístico, a partir de unos ajustes sencillos que admitirían diversas alternativas; los de *Agostini* y *Engelman*, suponen una mejora en la explicación del fenómeno de la acumulación y, en el segundo caso, la extensión a otros medios; y el de *Hofmans*, que se complementa con sus fórmulas de la audiencia neta y la audiencia neta acumulada, llegando a una explicación global de gran valor teórico y con validez empírica.

1. Modelo Logarítmico

Este modelo se basa en la estimación de la cobertura acumulada a partir de una ecuación logarítmica, de la siguiente forma:

$$C_{n,i} = a + b \cdot \log(n)$$

donde $C_{n,i}$ es la cobertura acumulada n inserciones en la revista i ; a y b los parámetros estimados para cada revista; y n el número de inserciones (bajo la hipótesis de una inserción por ejemplar).

A partir de dos bases de datos¹⁰⁶, Chandon aplicó este modelo a cinco revistas, estudiando la acumulación hasta 6 inserciones. La regresión logarítmica se ajustó para cada revista por separado, dando lugar a modelos siempre con un R^2 mayor del 99'5%, menos en una revista que rondaba el 99%. La estimación de la audiencia acumulada era muy precisa en todos los casos, para cualquier revista y cualquier número de inserciones. La aplicación general de este método exige conocer la audiencia acumulada de un buen número de inserciones: como habitualmente no estaba disponible más que la audiencia de las dos primeras inserciones, Chandon planteó una modificación en el modelo de acumulación logarítmica. Conociendo el alcance de los dos primeros ejemplares, se plantea un sistema de ecuaciones, que sirve para estimar de los parámetros a y b :

$$\begin{cases} C_1 = a \\ C_2 = a + b \cdot \log(2) \end{cases} \Rightarrow \begin{cases} \hat{a} = C_1 \\ \hat{b} = (C_2 - C_1) / \log(2) \end{cases}$$

¹⁰⁶ ALFRED POLITZ RESEARCH, INC. "Life Study of Consumer Expenditures". 7 Volumes. Time, Inc. New York. 1958.

ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: Their Accumulative and Repeat Audiences". Time, Inc. New York. 1953.

de manera que la cobertura acumulada de n inserciones tomaría esta nueva expresión:

$$C_{n,i} = C_1 + \frac{C_2 - C_1}{\log(2)} \cdot \log(n)$$

Los valores de \hat{a} y \hat{b} , estimados a partir de los dos primeros ejemplares (o inserciones, en nuestro planteamiento), no diferían mucho de los parámetros a y b estimados a partir de todos los ejemplares. Esas diferencias producían, en cuatro de las cinco revistas, ligeras y crecientes subestimaciones. Así este método permite conocer la acumulación de audiencia de un soporte a partir de su audiencia con una y dos inserciones.

Teóricamente plantea un problema: es evidente que la cobertura acumulada está acotada inferiormente en 0 y superiormente en 1. Sin embargo, dado el recorrido de una función logarítmica, la cobertura acumulada oscilaría en el intervalo $(-\infty, +\infty)$: cuando $n = 0$, la cobertura acumulada sería menos infinito; por el contrario, cuanto mayor sea n , en vez de tender asintóticamente a uno, tenderá a infinito; despejando se podría obtener el número de inserción a partir del cual la cobertura se hace mayor que uno: $n = 10^{(1-a)/b}$. En la práctica este problema no tiene trascendencia, porque el número de inserciones ni es nulo ni es muy elevado, dadas las características de las campañas.

2. Modelo Exponencial

Este modelo permite hallar la cobertura acumulada de n inserciones en un soporte, a partir de las coberturas acumuladas de las tres primeras inserciones, una más que en el modelo Logarítmico. Teóricamente es más correcto, ya que plantea que el soporte tiene una cobertura máxima (\bar{C}) que no puede superarse: la función es, por tanto, asintótica en esa cobertura máxima y puede expresarse:

$$C_n = \bar{C} - a \cdot b^n \quad 0 < b < 1$$

Los parámetros se estiman a partir de la coberturas acumuladas de las tres primeras inserciones:

$$\hat{\bar{C}} = \frac{(C_2)^2 - C_1 \cdot C_3}{2 \cdot C_2 - C_1 - C_3} = C_1 + a \cdot b$$

$$\hat{a} = \frac{(\bar{C} - C_1)^2}{\bar{C} - C_2} \quad \hat{b} = \frac{\bar{C} - C_2}{\bar{C} - C_1}$$

En cuanto a la lógica matemática del modelo, se mantiene la misma incoherencia cuando el número de inserciones es nulo: en ese caso la cobertura acumulada sería igual a la cobertura máxima, lo cual no tiene sentido. Desde un punto de vista práctico, estas consideraciones deben obviarse.

El modelo de Acumulación Exponencial también admite otra formulación:

$$C_n = C_1 - (\bar{C} - C_1) \cdot (1 - e^{c(n-1)}) \quad c < 1$$

determinándose los parámetros de esta forma:

$$\hat{c} = \ln \frac{C_1 - C_3}{C_1 - C_2 - 1} \quad \hat{\bar{C}} = C_1 + (C_2 - C_1) \cdot (1 - e^{\hat{c}})$$

Cuando Chandon¹⁰⁷ aplicó este modelo sobre las referidas encuestas de Politz (1953 y 1958), encontró subestimaciones sistemáticas y crecientes de la cobertura acumulada. El error estándar había aumentado en comparación con el modelo logarítmico. En varias ocasiones la cobertura acumulada con seis inserciones era mayor que la cobertura acumulada máxima. Esta inconsistencia se debe, a mi juicio, a que esa cobertura acumulada máxima ha sido estimada con un margen de error, pues se hace sólo a partir de las tres primeras inserciones.

Como conclusión, Chandon señala que, aunque el modelo exponencial fuera más coherente lógicamente, es preferible el logarítmico porque nos permitirá estimar con menos errores, sin olvidar que requiere menos información inicial.

¹⁰⁷ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 138.

3. Modelo de Agostini

Agostini¹⁰⁸ propuso un modelo que estima el alcance acumulado del n-ésimo ejemplar de una revista, a partir de los alcances del primero y de los dos primeros ejemplares. Las encuestas disponibles en ese momento eran escasas y caras. Así, los estudios de Politz, sobre cinco revistas, no consideraban más de seis inserciones: para ello, el de 1953¹⁰⁹ requirió una muestra superior a 7.000 individuos, entrevistados seis veces, con un intervalo de ocho semanas entre llamadas; y el de 1958¹¹⁰ necesitó una muestra de más de 15.000 hogares, entrevistados cuatro veces en intervalos de cuatro semanas.

Inicialmente asume que la acumulación de audiencia es un proceso aleatorio, es decir, que no existe relación entre los lectores de una y otra inserción de la misma revista. Siendo C_i la cobertura acumulada de i inserciones o ejemplares, y $1-C_i$ su no cobertura acumulada, se supone que cada nueva inserción alcanza a un porcentaje constante (P) de la no cobertura acumulada correspondiente:

$$\begin{aligned} C_2 &= C_1 + P \cdot (1 - C_1) \Rightarrow C_2 - C_1 = P \cdot (1 - C_1) \\ C_3 &= C_2 + P \cdot (1 - C_2) \Rightarrow C_3 - C_2 = P \cdot (1 - C_2) \\ &\dots \\ C_n &= C_{n-1} + P \cdot (1 - C_{n-1}) \Rightarrow C_n - C_{n-1} = P \cdot (1 - C_{n-1}) \end{aligned}$$

Podríamos volver a escribir dichas expresiones como sigue:

$$P = \frac{C_2 - C_1}{1 - C_1} = \frac{C_3 - C_2}{1 - C_2} = \dots = \frac{C_n - C_{n-1}}{1 - C_{n-1}}$$

¹⁰⁸ AGOSTINI, Jean-Michel. "Analysis of Magazine Accumulative Audience". *Journal of Advertising Research*. Vol. 2. Nº 4. Diciembre 1962. Págs. 24-7.

¹⁰⁹ ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: Their Accumulative and Repeat Audiences". Time, Inc. New York. 1953.

¹¹⁰ ALFRED POLITZ RESEARCH, INC. "Life Study of Consumer Expenditures". 7 Volumes. Time, Inc. New York. 1958.

Si cada nueva inserción añade ese porcentaje constante (P) de la no cobertura acumulada, también se podría aplicar a la cobertura acumulada de la primera inserción:

$$C_1 = C_0 + P \cdot (1 - C_0) = P$$

Sustituyendo en la cobertura acumulada de dos inserciones:

$$C_2 = C_1 + C_1 \cdot (1 - C_1) \Rightarrow 1 - C_2 = (1 - C_1)^2$$

Y generalizando para n inserciones:

$$1 - C_n = (1 - C_1)^n$$

Aplicando este modelo a las cinco revistas del estudio de Politz (1953 y 1958), encontró que P no era constante ni igual a C_1 , sino que en todos los casos iba decreciendo regularmente al aumentar el número de inserciones. La acumulación no es entonces un proceso aleatorio: estar expuesto a una inserción y estar expuesto a otra no son dos fenómenos independientes, sino que están correlacionados.

Agostini encontró que la relación entre las variables P y n , se puede explicar a través de la siguiente función logarítmica:

$$\log P = b - a \cdot \log n$$

Cuando n vale 1, P es igual a C_1 , y se cumplirá que $b = \log C_1$. Luego la fórmula se puede volver a escribir así:

$$\log P = \log C_1 - a \cdot \log n$$

Tomando la inversa en ambos miembros, e igualando al valor de P , obtenemos:

$$P = \frac{C_1}{n^a} = \frac{C_n - C_{n-1}}{1 - C_{n-1}}$$

Y operando se llega a la relación entre las coberturas acumuladas de n y $n-1$ inserciones:

$$1 - \frac{C_1}{n^a} = 1 - \frac{C_n - C_{n-1}}{1 - C_{n-1}} \Rightarrow 1 - C_n = (1 - C_{n-1}) \cdot \left(1 - \frac{C_1}{n^a}\right)$$

La fórmula general propuesta para la cobertura acumulada de n inserciones es:

$$1 - C_n = (1 - C_1) \cdot \left(1 - \frac{C_1}{2^a}\right) \cdot \dots \cdot \left(1 - \frac{C_1}{n^a}\right)$$

Para obtener el valor de a , se despeja en esta fórmula suponiendo que $n=2$:

$$2^a = \frac{C_1 \cdot (1 - C_1)}{C_2 - C_1} \Rightarrow a = \frac{1}{\log 2} \cdot \log \frac{C_1 \cdot (1 - C_1)}{C_2 - C_1}$$

Si se admitiera que la acumulación es un proceso aleatorio, se obtendría una ecuación tal como la inicial:

$$a = \frac{1}{\log 2} \cdot \log 1 = 0 \Rightarrow$$

$$1 - C_n = (1 - C_1) \cdot \left(1 - \frac{C_1}{2^0}\right) \cdot \dots \cdot \left(1 - \frac{C_1}{n^0}\right) = (1 - C_1)^n$$

Agostini, a partir de su método no estocástico, estimó la cobertura acumulada de las 5 revistas del estudio de Politz, comparando los resultados. Hasta 6 inserciones la desviación media era del 0,65%, con una desviación máxima del 2%. De 7 a 13 inserciones, no siendo los valores de Politz reales sino estimados, presentaron una desviación ligeramente mayor: por ejemplo, para 12 inserciones, la desviación media era del 1,35%, con una desviación máxima del 3,4%.

El método propuesto estima satisfactoriamente, según el autor, la cobertura acumulada dentro de los distintos públicos objetivo. Sin embargo, igual que el modelo

matemático de Politz, para *targets* complementarios como hombres y mujeres, el método da dos estimaciones cuya suma no es necesariamente igual a la estimación obtenida para toda la población.

Este modelo también puede calcular la cobertura neta acumulada (objeto de estudio de la cuarta parte), siempre que la campaña conste del mismo número de inserciones en distintos soportes. Todo ello a partir de C_1 y C_2 , siendo C_1 la cobertura neta de una inserción en cada soporte y C_2 la cobertura neta acumulada de dos inserciones en cada soporte. Probando para las revistas de Politz, Agostini concluyó que las estimaciones eran bastante precisas. No obstante, lo pudo verificar en muy pocos casos.

Como conclusión, Agostini señala que la acumulación de audiencia sigue esta ley matemática: cada nueva inserción añade a la cobertura ya alcanzada una fracción decreciente de la población no cubierta; esta fracción decrece precisamente $n^{-\alpha}$, donde n es el número de inserciones y α es un parámetro inherente a cada revista. Esta ley empírica debería ser testada para otros países y otros medios.

4. Modelo de Engelman

Engelman¹¹¹ presentó un modelo de acumulación en Revistas y Televisión, a partir del alcance acumulado por las dos primeras inserciones. Requería, por tanto, menos información que las seis primeras inserciones del modelo de Politz (posteriormente comentado) y que las cuatro necesitadas por el estudio de Audits & Surveys¹¹². La precisa fórmula de Agostini empleaba también dos inserciones, pero resultaba inadecuada para Young & Rubicam, la Agencia donde trabajaba Engelman.

Para eliminar la diferencias de audiencias entre soportes, utilizó el concepto de frecuencia (número de impactos dividido por número de impactados). Estableció la hipótesis de partida de su modelo: existe una curva ideal de acumulación, o al menos una curva ideal de incrementos de frecuencia, basada en el porcentaje de incremento de la frecuencia en la segunda inserción y en la frecuencia relativa de los incrementos para cada inserción adicional. Así se expresa su fórmula:

$$A_{n,i} = \frac{A_i \cdot n}{1 + a \cdot b}$$

siendo $A_{n,i}$ el alcance acumulado de la revista i para la inserción n ; A_i el alcance de su primera inserción; a el incremento de frecuencia en la segunda inserción; y b la acumulación de los incrementos relativos de frecuencia para n inserciones.

El alcance acumulado será, por tanto, la audiencia bruta ($A_i \cdot n$) debidamente corregida por la frecuencia media ($a \cdot b$). El valor de a es constante y propio de cada revista. Los valores de b deberían ser los mismos para todas las revistas, aunque varían para cada público objetivo. Young & Rubicam había determinado empíricamente los valores de b , como aparecen en la **Tabla 3.2**:

¹¹¹ ENGELMAN, Fred L. "An Empirical Formula for Audience Accumulation". *Journal of Advertising Research*. Vol. 5. Nº 2. Junio 1965. Págs 21-8.

¹¹² AUDITS & SURVEYS COMPANY, INC. "Look Audience Study, 1961". New York: Cowles Magazines and Broadcasting, Inc., 1961.

Nº de Inserciones	Valor de b para Hogares	Valor de b para Individuos
3	1,91	1,92
4	2,78	2,79
5	3,62	3,62
6	4,44	4,41
7	5,26	5,19
8	6,08	5,96
9	6,90	6,72
10	7,72	7,47
11	8,54	8,21
12	9,36	8,95
13	10,20	9,70

Tabla 3.2: Valores de b según el nº de inserciones y el público referido.

El alcance acumulado por las revistas de Politz y Audits & Surveys, fue estimada con precisión por este método, obteniéndose una diferencia media del 1%, en valor absoluto, y para todas las alternativas de inserciones y de públicos.

Estableció un método para determinar a cuando no había datos disponibles, basándose en la hipótesis de que la acumulación primaria de la audiencia es virtualmente idéntica para todas las publicaciones. Dicho procedimiento, como él mismo señalaba, carecía de una adecuada justificación empírica y debía revisarse cuando existieran más datos. Por otra parte, como a varía según el target considerado, proponía otro método de cálculo (tampoco adecuadamente justificado) cuando no estuviera disponible la frecuencia de la segunda inserción:

$$a_j = a \cdot \sqrt{\frac{C_{1,j}}{C_1}} \quad \forall C_{1,j} > C_1 \quad \text{ó bien} \quad a_j = a \cdot \sqrt[3]{\frac{C_{1,j}}{C_1}} \quad \forall C_{1,j} < C_1$$

siendo a_j el valor de a para el target j ; a ese valor para toda la población; $C_{1,j}$ la cobertura acumulada de una inserción en el target j ; y C_1 la cobertura acumulada de una inserción en toda la población.

Engelman analizó también la acumulación de la audiencia televisiva, referida a hogares, a partir de los datos de Nielsen Television Index¹¹³. Con su método consiguió estimaciones precisas, y dedujo los valores de b para el medio TV, como se muestran en la **Tabla 3.3**.

El modelo de Engelman debe circunscribirse a las fuentes de datos utilizadas por el autor. Al seguir un método inductivo, ha pretendido generalizar a partir de lo particular, de manera que la conclusión inducida está limitada a los fuentes de datos. Me parece que esto explica las divergencias con los datos de Simmons¹¹⁴, recogidas en el citado artículo de Engelman.

Nº de Inserciones	Valor de b para Inserciones Diurnas	Valor de b para Inserciones Nocturnas
3	1,87	1,89
4	2,71	2,74
5	3,38	3,55
6	3,97	4,32
7	4,73	5,07
8	5,30	5,80
9	5,74	6,53
10	6,17	7,26
11	6,59	7,99
12	7,10	8,72
13	7,67	9,47
14	8,22	-
15	8,71	-
16	9,15	-
17	9,59	-
18	10,06	-
19	10,56	-
20	11,03	-

Tabla 3.3: Valores de b según el nº de inserciones en TV y su horario.

¹¹³ NIELSEN, A. C., COMPANY. "Nielsen Television Index". 1959-64.

¹¹⁴ SIMMONS, W. R. AND ASSOCIATES RESEARCH, INC. "Mass Markets and the Media Reaching Them, 1964". New York: W. R. Simmons and Associates Research, Inc, 1964.

SIMMONS, W. R. AND ASSOCIATES RESEARCH, INC. "Selective Markets and the Media Reaching Them, 1964". New York: W. R. Simmons and Associates Research, Inc, 1964.

5. Modelo de Hofmans

Hofmans¹¹⁵ llegó a una fórmula para estimar el alcance acumulado por n inserciones en un soporte, con una estructura similar a su fórmula del alcance neto de una inserción en n soportes. Necesitando sólo los alcances de las inserciones primera y segunda (C_1 y C_2), el alcance acumulado de n inserciones se expresa en los siguientes términos:

$$C_n = \frac{(nC_1)^2}{nC_1 + \frac{n(n-1)}{2} \cdot kd}$$

$$\text{siendo } d \approx 2C_1 - C_2 \quad \text{y} \quad kd = \frac{2C_1 \cdot d}{2C_1 - d}$$

Esta fórmula está considerando constante la duplicación entre dos inserciones (d) y, por tanto, también será constante $k \cdot d$. Con esta hipótesis, los resultados se compararon con diferentes estudios, observando que los valores de C_n eran subestimados, con diferencias crecientes cuando n era mayor. Hofmans concluía que uno de los valores incluidos en la fórmula debería variar con n .

Primeramente investigó la **constancia de la duplicación (d)**. Si se admite la constancia de la audiencia de la primera inserción (audiencia media del soporte) debe cumplirse:

$$C_1 = C_1,$$

$$C_2 = 2C_1 - d_2,$$

$$C_3 = 3C_1 - 3d_3 + t_3,$$

$$C_4 = 4C_1 - 6d_4 + 4t_4 - q_4 \quad \dots$$

¹¹⁵ HOFMANS, PIERRE. "Measuring the Cumulative Net Coverage of Any Combination of Media". *Journal of Marketing Research*. Agosto de 1966. Vol. 3. Págs. 271-75.

En esas relaciones, d_2, d_3, d_4, \dots , son los diferentes valores de la duplicación; t_3, t_4, \dots , son los diferentes valores de la triplicación; q_4, \dots , son los diferentes valores de la cuadruplicación; etc. Si se admite que la duplicación permanece constante:

$$2C_1 - C_2 = d_2 = d_3 = d_4 = \dots$$

Sustituyendo en el alcance acumulado de la tercera inserción, se obtiene el valor constante de las triplicaciones:

$$C_3 = 3C_1 - 3(2C_1 - C_2) + t_3 \Leftrightarrow$$

$$3C_1 - 3C_2 + C_3 = t_3 = t_4 = \dots$$

Del mismo modo para la cuarta inserción:

$$C_4 = 4C_1 - 6(2C_1 - C_2) + 4(3C_1 - 3C_2 + C_3) - q_4 \Leftrightarrow$$

$$4C_1 - 6C_2 + 4C_3 - C_4 = q_4 = \dots$$

Igualmente se podría determinar el caso extremo de los que contactan con cada una de las n inserciones:

$$nC_1 - \frac{n(n-1)}{2!}C_2 + \frac{n(n-1)(n-2)}{3!}C_3 - \dots \pm nC_{n-1} \pm C_n$$

Esta fórmula fue verificada para el estudio de Politz¹¹⁶. Por tanto, al aceptarse que la duplicación d es constante, es preciso afirmar la **variación de k** . Suponiendo que k toma los valores $k_1, k_2, k_3, \dots, k_n$, según el número de inserciones:

$$k_n = \frac{2C_1}{d(n-1)} \cdot \frac{nC_1 - C_n}{C_n} \Leftrightarrow \frac{nC_1 - C_n}{C_n} = \frac{k_n \cdot d(n-1)}{2C_1} \quad [1]$$

Para dos inserciones tendremos:

¹¹⁶ ALFRED POLITZ RESEARCH, INC. "Life Study of Consumer Expenditures". New Time, Inc. 1959.

$$k_2 = \frac{2C_1}{d} \cdot \frac{2C_1 - C_2}{C_2} \Leftrightarrow \frac{2C_1 - C_2}{C_2} = \frac{k_2 \cdot d}{2C_1} \quad [2]$$

En dos estudios¹¹⁷ se verificó, para siete revistas, la siguiente relación:

$$\log \frac{C_n}{nC_1 - C_n} = b - a \cdot \log (n-1)$$

Cuando $n = 2$, se cumple que $b = \log \frac{C_2}{2C_1 - C_2}$. Sustituyendo el valor de b

en la expresión anterior:

$$\log \frac{C_n}{nC_1 - C_n} = \log \frac{C_2}{2C_1 - C_2} - a \cdot \log (n-1) \Leftrightarrow$$

$$\frac{nC_1 - C_n}{C_n} = \frac{2C_1 - C_2}{C_2} \cdot (n-1)^a \quad [3]$$

Sustituyendo los valores de las ecuaciones [1] y [2]:

$$\frac{k_n \cdot d(n-1)}{2C_1} = \frac{k_2 \cdot d}{2C_1} \cdot (n-1)^a \Leftrightarrow k_n \cdot (n-1) = k_2 \cdot (n-1)^a$$

Por tanto, la fórmula del alcance acumulado de n inserciones vendrá finalmente determinada así:

$$C_n = \frac{(nC_1)^2}{nC_1 + \frac{n(n-1)^a}{2} \cdot kd} \quad [4]$$

De ahí se puede extraer el valor de kd :

¹¹⁷ Véase:

ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: A Study of Duplication". Time, Inc. New York. 1953.

SOFRES (Société Française d'Enquêtes par Sondage): "Etude des 3 Supports". Paris. 1965.

$$k_d = \frac{2C_1 \cdot (nC_1 - C_n)}{C_n \cdot (n - 1)^a}$$

Hofmans señala que la fórmula [4] puede calcular el alcance acumulado en cualquier medio impreso, radio, televisión o cine. Para usarla es necesario y suficiente conocer el valor de C_1 y dos diferentes valores C_i , digamos C_s y C_t (con $C_s > C_t$).

El valor de a puede ser despejado de la ecuación [3], para s y t inserciones:

$$\left. \begin{aligned} (s-1)^a &= \frac{sC_1 - C_s}{C_s} \cdot \frac{C_2}{2C_1 - C_2} \\ (t-1)^a &= \frac{tC_1 - C_t}{C_t} \cdot \frac{C_2}{2C_1 - C_2} \end{aligned} \right\} \Rightarrow a = \frac{\log \frac{C_t \cdot (sC_1 - C_s)}{C_s \cdot (tC_1 - C_t)}}{\log \frac{s-1}{t-1}}$$

Hofmans calculó el valor de a para $s=4$ y $t=2$, y estimó el alcance acumulado de diversas revistas, con distinto número de inserciones y referida a varios *target*. Los resultados fueron satisfactorios, pues los errores alcanzados eran siempre menores del 2,4% y siempre por debajo de los errores muestrales. Se concluía además que a es una constante propia de cada revista, independiente del universo considerado. Además, comparando el valor de a de cinco revistas para distintos años de los estudios de Politz, parece concluirse que a es independiente también del estudio.

Hofmans apunta finalmente que sería muy conveniente la **constancia de la duplicación**, pues simplifica la información necesaria. Para contrastarlo emplea el parámetro definido por Durand¹¹⁸:

$$\lambda = \frac{\log (A_1 \cdot A_2)}{\log D_{12}}$$

¹¹⁸ DURAND, JACQUES. "L'Analyse Statistique de la Duplication". Institut de Recherches et d'Etudes Publicitaires (IREP). N° 11. París. Noviembre 1964.

siendo A_1 y A_2 las audiencias de dos soportes, y D_{12} su audiencia duplicada. Para la acumulación de audiencia de un mismo soporte, las dos inserciones tienen la misma audiencia, y el parámetro quedará definido así:

$$\lambda = \frac{2 \cdot \log C_1}{\log d}$$

Si este coeficiente de duplicación fuera estable, podríamos utilizarlo para predecir el valor de la duplicación d , despejando de la fórmula anterior. Se podría calcular $k \cdot d$ y, conociendo el valor de α para una revista determinada, se podría estimar su alcance acumulado para n inserciones.

Cuando Chandon¹¹⁹ testó la fórmula [4], encontró unos resultados menos satisfactorios que los modelos de acumulación logarítmica y de Agostini. Aunque estuviera mejor fundamentado teóricamente, el modelo de Hofmans producía sistemáticamente sobrestimaciones. Además requiere el conocimiento del alcance acumulado de tres inserciones.

¹¹⁹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 147-9.

B) MODELOS ESTOCÁSTICOS

Los Modelos Estocásticos de Acumulación, a diferencia de los Determinísticos, asumen que la exposición es un fenómeno probabilístico. Se basan en una serie de hipótesis verificables acerca del comportamiento individual del lector, espectador u oyente. Permiten estimar el alcance acumulado y la distribución de contactos, a diferencia de los Modelos Ad-Hoc que se limitaban al alcance acumulado.

Mantienen la hipótesis fundamental de que los individuos de la población objetivo poseen una cierta probabilidad personal de exposición que explica su comportamiento ante el soporte¹²⁰. En el modelo más completo, esa probabilidad debería depender del individuo, del soporte, de la inserción, de las exposiciones pasadas e incluso de otras probabilidades individuales de exposición. Sin embargo, un proceso estocástico tan complejo sería inmanejable. Se puede simplificar manteniendo la *estacionariedad* (que las cosas que hayan ocurrido en el pasado permanezcan estacionarias) y aceptando o rechazando la *homogeneidad* (que los individuos se comporten de la misma forma que la población en su conjunto). La hipótesis de *estacionariedad* es de dos tipos:

- estacionariedad a través de las *inserciones*: la probabilidad individual de exposición a una inserción determinada es independiente de las exposiciones pasadas e independiente de la particular naturaleza de tiempo y contenido de la inserción.

- estacionariedad a través de los *individuos*: la probabilidad de exposición de un individuo *i* a la inserción *n*, es independiente de la probabilidad de exposición del individuo *j* para la misma inserción. En otros términos, esta hipótesis excluye la posibilidad de cualquier comportamiento individual no independiente.

¹²⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 150.

La hipótesis de la *homogeneidad* supone que las probabilidades de exposición son iguales para todos los individuos. Su antagónica heterogeneidad puede tomar tres formas:

- heterogeneidad de los *segmentos de población*: cada *target* se caracteriza por una singular probabilidad de exposición; la homogeneidad se da dentro de los individuos del segmento.

- heterogeneidad de los *individuos*: cada individuo se caracteriza por tener su personal probabilidad de exposición. Dicho de otra manera, las probabilidades individuales de exposición siguen alguna distribución independiente del *target* considerado.

- heterogeneidad *mixta*: cada segmento de población se caracteriza por tener una particular distribución de probabilidades individuales de exposición.

La combinación de las alternativas de estacionariedad y homogeneidad, diferencian los modelos de acumulación de audiencia. Por no extenderme demasiado, a continuación sólo se describen tres: Binomial, Heterogeneidad Lineal y Beta Binomial; otros pueden encontrarse en las referencias: Broadbent¹²¹, Morgensztern¹²² y Binomial negativa¹²³.

¹²¹ BROADBENT, Simon. "Regularity of Reading". *Journal of Marketing Research*. Agosto 1964. Págs. 50-8.

¹²² MORGENSZTERN, Armand. "Frequentation de la Presse, Generalisation, Effet Publicitaire". *IREP, Etude* Nº 19. París. Abril 1970.

¹²³ GREENE, Jerome D. and STOCK, J. Stevens. "A Rate-Frequency Model of Behavior". *Journal of Advertising Research*. Vol. 11. Nº 4. Agosto 1971. Págs. 9-19.

1. Modelo Binomial

El Modelo Binomial es el más sencillo de los modelos estocásticos de acumulación, pues se basa en las hipótesis menos restrictivas:

- La población es homogénea: todos los individuos tienen la misma probabilidad de exposición a una inserción determinada de un soporte dado. Siendo $I=\{1, 2, \dots, i, \dots, I\}$ un grupo de individuos y $N=\{1, 2, \dots, n, \dots, N\}$ una serie de inserciones, las probabilidades individuales cumplen:

$$p_{1,n} = p_{2,n} = \dots = p_{i,n} = \dots = p_n$$

- Las probabilidades de exposición son estacionarias desde una inserción a la siguiente. Así, todas las inserciones en un soporte específico son consideradas equiprobables, y se expresan:

$$p_{1,1} = p_{1,2} = \dots = p_{1,n} = \dots = p_1$$

Resumiendo, la probabilidad de que un individuo cualquiera esté expuesto a cualquiera de las inserciones es constante para todos los individuos y para todas las inserciones. Notaremos dicha probabilidad como p , y será el cociente entre la audiencia media del soporte y la población objetivo.

Dadas las hipótesis del modelo, la exposición es una variable de Bernouilli. Por lo tanto, la probabilidad de que un individuo esté expuesto exactamente a k inserciones entre n posibles, viene determinada así:

$$P(k/n, p) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad 0 \leq k \leq n$$

y el número esperado de individuos expuestos exactamente a k inserciones entre n posibles:

$$C_k = P(k/n, p) \cdot P$$

La probabilidad de que un individuo esté expuesto al menos a k inserciones entre las n posibles, se expresa:

$$P(x \geq k/n, p) = \sum_{x=k}^n \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Para el caso particular del alcance acumulado en n inserciones, la probabilidad de haber recibido al menos una inserción es:

$$P(x \geq 1/n, p) = 1 - P(x = 0/n, p) = 1 - (1-p)^n$$

y, dada la homogeneidad de la población, el alcance acumulado de n inserciones vale:

$$C_n = [1 - (1-p)^n] \cdot P$$

El modelo binomial sobrestima sistemática y considerablemente el alcance acumulado en el estudio de Chandon¹²⁴. Ello se debe a que las hipótesis asumidas son muy irrealistas: por ejemplo, los datos avalan que la persona, que contacta con una inserción de un soporte, tiene mayor probabilidad de contactar con otra inserción de ese soporte, y no la misma probabilidad postulada por el modelo binomial.

¹²⁴ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 154-6.

2. Modelo de heterogeneidad lineal

Se basa en la heterogeneidad de la población, ya que cada individuo tiene su propia probabilidad de exposición a un soporte dado; también esas probabilidades son estacionarias entre los distintos números y los distintos individuos. Las probabilidades individuales pueden tomar cualquier valor entre 0 y 1, así que la distribución sobre toda la población es continua. El alcance acumulado y la distribución de contactos se va estimar a partir de $f(p_i)$, que es el número de individuos asociados con la probabilidad p_i . Según sea la forma de $f(p_i)$, así se estimará la distribución de contactos. Como referencia Chandon, Politz explicó la relación que vincula $f(p_i)$ con E_k^n (numero de individuos expuestos a k de las n inserciones de un soporte) y con R_n (número de individuos expuestos al menos a una de las n inserciones en un soporte, o alcance acumulado de n inserciones). La población objetivo (P) es la suma de todos los valores tomados por $f(p_i)$ o la suma de la distribución de contactos:

$$P = \int_0^1 f(p_i) dp = \sum_{k=0}^n E_k^n = R_n + E_0^n$$

Como se acepta la estacionariedad, la exposición o no exposición del individuo i se puede considerar una variable aleatoria discreta, que toma valores uno y cero, respectivamente con probabilidades p_i y q_i ($1-p_i$). La probabilidad de recibir k entre n exposiciones está determinada por la distribución binomial:

$$P(k/n, p_i) = \binom{n}{k} p_i^k (1-p_i)^{n-k} \quad 0 \leq k \leq n$$

Existen $f(p_i)$ individuos con esa probabilidad de recibir k entre n contactos. Sumando todos los individuos, para todas las posibles probabilidades, se puede obtener el número esperado de individuos que reciben k entre n exposiciones:

$$E_k^n = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p) dp$$

Chandon encontró una relación entre E_k^n y E_k^{n-1} , que se cumple para cualquier forma de $f(p)$ y permite conocer la frecuencia de cualquier combinación menor de n números. Multiplicando y dividiendo la anterior expresión por $(1-p)$:

$$E_k^n = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k-1} (1-p) f(p) dp = \int_0^1 \binom{n}{k} [p^k (1-p)^{n-k-1} - p^{k+1} (1-p)^{n-k-1}] f(p) dp \Rightarrow$$

$$\frac{E_k^n}{\binom{n}{k}} = \int_0^1 p^k (1-p)^{n-k-1} f(p) dp - \int_0^1 p^{k+1} (1-p)^{n-k-1} f(p) dp \Rightarrow$$

$$\frac{E_k^n}{\binom{n}{k}} = \frac{E_k^{n-1}}{\binom{n-1}{k}} - \frac{E_{k+1}^n}{\binom{n}{k+1}} \Rightarrow E_k^{n-1} = \frac{n-k}{n} E_k^n + \frac{k+1}{n} E_{k+1}^n$$

Es fácil derivar la misma relación para el alcance acumulado: $R_{n-1} = R_n - E_1^n / n$. Se puede generalizar también para derivar la distribución de frecuencia de $n-t$ números a partir de la distribución de frecuencia de n :

$$E_k^{n-t} = \sum_{i=0}^t \frac{\binom{t}{i} \binom{n-t}{k}}{\binom{n}{k+i}} E_{k+i}^n$$

Politz¹²⁵ asumió que $f(p)$ debería aproximarse satisfactoriamente a un polinomio de tercer grado:

$$f(p) = a + b \cdot p + c \cdot p^2 + d \cdot p^3$$

Como depende de cuatro parámetros (a , b , c y d) es necesario conocer al menos el alcance acumulado por cuatro números. De hecho Politz estimó esos parámetros a partir de las acumulaciones de seis números, en las revistas de su estudio. Teniendo más ecuaciones que incógnitas, obtuvo la mejor estimación por mínimos cuadrados.

¹²⁵ ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: Their Accumulative and Repeat Audiences". Time, Inc. New York. 1953.

Algunos¹²⁶ han sugerido que una función cuadrática debería ser suficiente para estimar $f(p)$, pero todavía exigiría conocer la acumulación de audiencia de tres números.

Como señala Chandon, la estimación de $f(p)$ a partir de una línea recta supondría desvirtuar la heterogeneidad de la población. Por otra parte, es el único ajuste posible si sólo se dispone del alcance acumulado por dos números; tomaría esta forma:

$$f(p) = a + b \cdot p$$

Sustituyendo este **modelo de heterogeneidad lineal** en el número de individuos que reciben k entre n exposiciones:

$$E_k^n = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} (a + bp) dp = \binom{n}{k} \left[a \int_0^1 p^k (1-p)^{n-k} dp + b \int_0^1 p^{k+1} (1-p)^{n-k} dp \right]$$

Sustituyendo las integrales por las correspondientes funciones Beta:

$$E_k^n = \binom{n}{k} [a \cdot \beta(k+1, n-k+1) + b \cdot \beta(k+2, n-k+1)]$$

y, siendo n y k números enteros, la función Beta toma los siguientes valores:

$$\beta(k+1, n-k+1) = \frac{k! \cdot (n-k)!}{(n+1)!} \quad \beta(k+2, n-k+1) = \frac{(k+1)! \cdot (n-k)!}{(n+2)!}$$

Los parámetros a y b se pueden estimar a partir de E_0^2 y E_1^2 mediante el siguiente sistema de ecuaciones:

$$\begin{cases} E_1^2 = 2 \cdot [a \cdot \beta(2,2) + b \cdot \beta(3,2)] \\ E_0^2 = a \cdot \beta(1,3) + b \cdot \beta(2,3) \end{cases} \Rightarrow \begin{cases} \hat{a} = 6E_0^2 - 3E_1^2 \\ \hat{b} = 12(E_1^2 - E_0^2) \end{cases}$$

¹²⁶ CHARNES, A.; COOPER, N.W.; LEARNER, D.B.L.; and SNOW, E.F.: "Media Mixing for Distribution of Frequencies with β -like and Lognormal Distributions and Frequencies". Management Science Report N° 189. Carnegie-Mellon University. Pittsburg. Junio 1968.

Por su parte, el alcance acumulado se puede estimar como sigue:

$$R_n = P - [a \cdot \beta(1, n+1) + b \cdot \beta(2, n+1)]$$

Chandon estimó así la acumulación de audiencia de las diez revistas del estudio Politz. Encontró importantes diferencias con los alcances y las frecuencias reales. Además la *heterogeneidad lineal* producía resultados ilógicos: para todas las revistas, el parámetro b era mayor en valor absoluto que a , lo cual implica absurdamente que, para valores de p cercanos a 1, existe un número negativo de individuos con tales probabilidades. Lo mismo se puede observar en el siguiente ejemplo.

Ejemplo

Supongamos una revista cuyo primer número alcanza al 25% de la población objetivo, y el segundo número llega hasta el 30%; además el 20% de la población ha estado expuesto simultáneamente a ambos números. Vamos a estimar el alcance acumulado y la distribución de contactos de seis números de esa revista (todas las magnitudes están en porcentaje sobre la población objetivo):

$$E_0^2 = 1 - R_2 = 1 - 0.3 = 0.7 \quad E_1^2 = R_2 - E_2^2 = 0.3 - 0.2 = 0.1$$

$$\begin{cases} \hat{a} = 6E_0^2 - 3E_1^2 = 6 \cdot 0.7 - 3 \cdot 0.1 = 3.9 \\ \hat{b} = 12(E_1^2 - E_0^2) = 12(0.1 - 0.7) = -7.2 \end{cases}$$

La distribución de contactos se estima en la **Tabla 3.4**, recogiendo el proceso paso a paso, para facilitar su comprensión. En la última columna, aparecen resultados absurdos en los contactos superiores, debido a la inconsistencia del modelo de heterogeneidad lineal.

k	$\binom{6}{k}$	(1) $\beta(k+1, 6-k+1)$	(2) $\beta(k+2, 6-k+1)$	$E_k^6 = \binom{6}{k} [\hat{a}(1) + \hat{b}(2)]$
0	1	0,1429	0,0179	42,86%
1	6	0,0238	0,0060	30,00%
2	15	0,0095	0,0036	17,14%
3	20	0,0071	0,0036	4,29%
4	15	0,0095	0,0060	-8,57%
5	6	0,0238	0,0179	-21,43%
6	1	0,1429	0,1250	-34,29%

Tabla 3.4

El alcance acumulado de los seis números de la revista se calcula de este modo:

$$R_6 = 1 - [3 \cdot 9 \cdot \beta(1, 7) - 7 \cdot 2 \cdot \beta(2, 7)] = 1 - \left[3 \cdot 9 \frac{0! \cdot (6-0)!}{(6+1)!} - 7 \cdot 2 \frac{(0+1)! \cdot (6-0)!}{(6+2)!} \right] \approx 0,5714$$

o como el complementario de los que reciben 0 exposiciones: $R_6 = 1 - E_0^6 = 1 - 0,4286$.

Se observa la incoherencia de la distribución de contactos, ya que algunos resultados son negativos y porque el alcance es distinto de la suma de los que reciben más de un

impacto: $R_6 \neq \sum_{k=1}^6 E_k^6$.

3. Modelo Beta Binomial

Hyett¹²⁷ propuso que $f(p)$ se explicara a través de una distribución Beta. Tiene la ventaja de que está definida en el intervalo $[0,1]$ y, por tanto, no puede tomar valores negativos. Además el área bajo la distribución suma la unidad, evitando los resultados absurdos del anterior modelo. Pero no es la única distribución con estas propiedades; sin embargo, Chandon¹²⁸ señala que la distribución Beta presenta varias ventajas:

- es la conjugada de la distribución Binomial: esto significa que, cuando el parámetro de un proceso Bernouilli es tratado como una variable aleatoria p , la distribución de p , más conveniente *a priori* es la distribución Beta, porque cuando se toma una muestra del proceso, la posterior distribución será todavía una distribución Beta.

- la distribución Beta es extremadamente flexible: puede tomar forma de U o de campana, tener uno o dos puntos de inflexión, ser simétrica o asimétrica, desviada hacia la derecha o hacia la izquierda.

- depende sólo de dos parámetros, A y B, que toman distintos valores para cada soporte y que se pueden estimar a partir del alcance acumulado de dos números; luego permite extender a cualquier n .

- asumiendo que $f(p)$ se distribuye según una Beta, la distribución de contactos sigue una distribución simple, llamada Beta Binomial.

- la elección de la distribución Beta para $f(p)$ produce estimaciones muy precisas del alcance y la distribución de contactos; de hecho, desde 1964 se dejaron de publicar los alcances acumulados de más de dos números, excepto en el medio TV.

¹²⁷ HYETT, G. P.: "The Measurement of Readership". London School of Economics Paper. London. 1958.

¹²⁸ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 178-205.

Vista la conveniencia de la distribución Beta, se podría aplicar a la población de varias formas: la misma Beta para todos, o bien una distinta por cada segmento, con lo que se gana en rigor pero también en la exigencia de datos. Aquí sólo se recogerá el primer caso.

Asumiendo que $f(p)$ sigue una única distribución Beta con parámetros A y B:

$$f(p) = P(p / A, B) = \frac{(A + B - 1)!}{(A - 1)! \cdot (B - 1)!} p^{A-1} (1 - p)^{B-1}$$

Los parámetros A y B han de ser números reales positivos, pero no necesariamente enteros. Por eso se puede expresar en los términos de una función Gamma:

$$\frac{(A + B - 1)!}{(A - 1)! \cdot (B - 1)!} = \frac{\Gamma(A + B)}{\Gamma(A) \cdot \Gamma(B)} \quad \text{con } \Gamma(A) = (A - 1)! \cdot \Gamma(A - 1)$$

El número esperado de individuos que recibirán k de las n posibles exposiciones será entonces:

$$E_k^n = \binom{n}{k} \frac{\Gamma(A + B)}{\Gamma(A) \cdot \Gamma(B)} \int_0^1 p^k (1 - p)^{n-k} p^{A-1} (1 - p)^{B-1} dp \Rightarrow$$

$$E_k^n = \binom{n}{k} \frac{\Gamma(A + B)}{\Gamma(A) \cdot \Gamma(B)} \int_0^1 p^{A+k-1} (1 - p)^{n+B-k-1} dp \Rightarrow$$

$$E_k^n = \binom{n}{k} \frac{\Gamma(A + B) \cdot \Gamma(A + k) \cdot \Gamma(n + B - k)}{\Gamma(A) \cdot \Gamma(B) \cdot \Gamma(A + B + n)} \Rightarrow$$

$$E_k^n = \binom{n}{k} \frac{(A + k - 1) \cdot (A + k - 2) \cdot \dots \cdot A \cdot (B + n - k - 1) \cdot (B + n - k - 2) \cdot \dots \cdot B}{(A + B + n - 1) \cdot (A + B + n - 2) \cdot \dots \cdot (A + B)}$$

El alcance acumulado de n números tendría una expresión más sencilla:

$$R_n = 1 - E_0^n = 1 - \frac{(B+n-1) \cdot (B+n-2) \cdot \dots \cdot B}{(A+B+n-1) \cdot (A+B+n-2) \cdot \dots \cdot (A+B)}$$

Suponiendo que hubiera datos del alcance acumulado de dos números (no el caso del EGM), los parámetros A y B se podrían calcular resolviendo este sistema de ecuaciones:

$$\begin{cases} R_1 = E_1^1 = \frac{A}{A+B} \\ R_2 = E_1^2 + E_2^2 = \frac{(A+1)A + 2AB}{(A+B+1)(A+B)} \end{cases}$$

Resolviendo el sistema los parámetros estimados valen:

$$\begin{cases} \hat{A} = \frac{E_1^1(E_1^1 - E_2^2)}{E_2^2 - (E_1^1)^2} = \frac{R_1(R_2 - R_1)}{2R_1 - (R_1)^2 - R_2} \\ \hat{B} = \frac{\hat{A}(1 - E_1^1)}{E_1^1} = \frac{\hat{A}(1 - R_1)}{R_1} \end{cases}$$

Según este procedimiento, Chandon testó el modelo para las revistas del estudio Politz. Aunque el alcance se subestimaba ligeramente, la distribución Beta era el modelo estocástico que mejor estimaba la distribución de contactos. Rust y Klompmaker¹²⁹ propusieron algunas modificaciones en el modelo Beta Binomial, para adaptarlo a las características del medio TV, obteniendo estimaciones más precisas.

Cuando se conoce la acumulación de más de dos números, existen varios métodos para estimar los parámetros A y B. Chandon recoge cinco, siendo los dos últimos originales: momentos, media y cero, mínimos cuadrados, mínima chi-cuadrado y

¹²⁹ RUST, Roland T. y KLOMPMAKER, Jay E. "Improving the Estimation Procedure for the Beta Binomial TV Exposure Model". *Journal of Marketing Research*. Vol. 18. Noviembre de 1981. Págs. 442-8.

máxima probabilidad. Su contenido se aleja de los objetivos de esta tesis. En el siguiente ejemplo veremos la superioridad de la Beta Binomial sobre anteriores modelos.

Ejemplo

Sobre los datos del anterior ejemplo, $R_1=0'25$, $R_2=0'3$ y $E_2^2=0'2$, la estimación de los parámetros conduce a estos resultados:

$$\begin{cases} \hat{A} = \frac{0'25(0'3 - 0'25)}{2 \cdot 0'25 - (0'25)^2 - 0'3} \approx 0'09 \\ \hat{B} = \frac{0'09 \cdot (1 - 0'25)}{0'25} \approx 0'27 \end{cases}$$

La distribución de contactos esperada, expresada en porcentaje sobre el público objetivo, se calcula de esta manera:

$$E_0^6 = \binom{6}{0} \frac{5'27 \cdot 4'27 \cdot 3'27 \cdot 2'27 \cdot 1'27 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'6304$$

$$E_1^6 = \binom{6}{1} \frac{0'09 \cdot 4'27 \cdot 3'27 \cdot 2'27 \cdot 1'27 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'0652$$

$$E_2^6 = \binom{6}{2} \frac{1'09 \cdot 0'09 \cdot 3'27 \cdot 2'27 \cdot 1'27 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'0416$$

$$E_3^6 = \binom{6}{3} \frac{2'09 \cdot 1'09 \cdot 0'09 \cdot 2'27 \cdot 1'27 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'0355$$

$$E_4^6 = \binom{6}{4} \frac{3'09 \cdot 2'09 \cdot 1'09 \cdot 0'09 \cdot 1'27 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'0362$$

$$E_5^6 = \binom{6}{5} \frac{4'09 \cdot 3'09 \cdot 2'09 \cdot 1'09 \cdot 0'09 \cdot 0'27}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'0465$$

$$E_6^6 = \binom{6}{6} \frac{5'09 \cdot 4'09 \cdot 3'09 \cdot 2'09 \cdot 1'09 \cdot 0'09}{5'36 \cdot 4'36 \cdot 3'36 \cdot 2'36 \cdot 1'36 \cdot 0'36} \approx 0'1447$$

La revista tiene, por tanto, una distribución de contactos en forma de U. Los resultados son todos positivos y suman la unidad, debido a las propiedades de la distribución Beta. Finalmente, el alcance acumulado de los seis números es el complementario de los que no reciben ningún contacto:

$$R_6 = 1 - E_0^6 = 1 - 0'6304 = 0'3694.$$

C) MODELOS DE SIMULACIÓN

La simulación supuso un nuevo enfoque de la planificación que, a diferencia de los anteriores, no analiza la información de forma agregada sino individuo a individuo. Aunque el recuento individual parezca más largo, no requiere los cruces de audiencias y duplicaciones en cada *target*, como algunos modelos de fórmula aquí vistos. Sin embargo, para Chandon era más fácil estimar los parámetros de la distribución agregada que agregar las distribuciones estimadas individualmente. Mi opinión, que supongo coincide con la actual de Chandon, es que el enfoque de la simulación es más adecuado y, hoy en día, es muy asequible gracias a la potencia de los ordenadores; el enfoque agregado probablemente no debe restringirse a un "mal menor", cuando no se dispone del tratamiento individualizado, sino que puede simplificar el proceso de optimización. El estudio de la simulación se acomete con más profundidad en la **IV Parte**.

Es necesario analizar dos cuestiones previas: cómo se determinan las probabilidades individuales y a qué individuos se asignan. Después ya se puede simular el comportamiento del individuo, en función de las probabilidades de exposición que tiene asignadas. Para cada cuestión previa se han planteado diferentes soluciones en la literatura, exponiéndose a continuación.

Según Schreiber¹³⁰, un modelo de simulación debe cumplir dos criterios: reproducir fielmente los datos de origen, y generar una única estimación o solución. Además los resultados deben ser coherentes. En función de estas restricciones evalúa varios métodos de simulación, que se exponen a continuación.

¹³⁰ SCHREIBER, Robert J.: "Probability Assignments for the Simulation of Media Reach and Frequency". *Journal of Advertising Research*. Junio 1968. Vol. 8. Nº 2. Págs. 3-8.

1. Método actuarial

La probabilidad individual de exposición se mide por el hábito de exposición, calculándose como la proporción entre las exposiciones realizadas y las exposiciones posibles. Por ejemplo, si se analizan semanarios durante un periodo de cuatro semanas, los individuos se pueden exponer 4, 3, 2, 1 ó 0 veces; el método actuarial les asignará respectivamente una probabilidad de exposición de 1 (4/4), 0'75 (3/4), 0'5 (2/4), 0'25 (1/4) y 0 (0/4). El nombre, sugerido por Greene¹³¹, hace referencia a la observación puntual del comportamiento, como originalmente se usaba en el sector de los seguros. Las exposiciones realmente realizadas se pueden conocer mediante encuesta o panel, aunque habitualmente se asocia con la encuesta; en ese caso presenta algunos inconvenientes:

- Las probabilidades se restringen a las respuestas cerradas del cuestionario. Así en el ejemplo anterior, ningún individuo tendrá asignadas probabilidades como 0'37, 0'61,..., pudiendo ser esas las que realmente sigue su comportamiento.

- La corrección de la respuesta depende de la memoria del entrevistado; cuanto menor es el periodo de observación, la respuesta se hace más fiable, pero las probabilidades pueden tomar menos alternativas.

Determinadas las probabilidades de exposición por el método actuarial, la distribución binomial permite conocer la distribución de contactos tras n exposiciones. Dada una población, con grupos de F_i individuos con sus correspondientes p_i probabilidades de exposición, el número de individuos que han recibido k exposiciones entre n (E_k^n) se expresa:

$$E_k^n = \binom{n}{k} \cdot \sum_{i=0}^n F_i \cdot p_i^k \cdot (1 - p_i)^{n-k}$$

¹³¹ GREENE, Jerome D.: "Personal Media Probabilities". *Journal of Advertising Research*. Vol. 10. Octubre 1970. Pág. 13.

En la **Tabla 3.5** se recoge el ejemplo expuesto por Schreiber: sobre 120 millones de individuos, 61 tienen una probabilidad nula de contactar con el soporte X, 20 una probabilidad del 25%, 14 una probabilidad del 50% ... (cfr. tercera columna). Esas probabilidades corresponden a la proporción de números vistos durante cuatro semanas. En la última fila aparecen las sumas de todos los grupos. La distribución de contactos total es la suma de los contactados k veces en cada columna.

Millones individ. F_i	Exposi- ciones recibidas	Prob. de exposic. p_i	0 contactos	1 contacto	2 contactos	3 contactos	4 contactos
			$F_i \binom{4}{0} p_i^0 (1-p_i)^4$	$F_i \binom{4}{1} p_i^1 (1-p_i)^3$	$F_i \binom{4}{2} p_i^2 (1-p_i)^2$	$F_i \binom{4}{3} p_i^3 (1-p_i)^1$	$F_i \binom{4}{4} p_i^4 (1-p_i)^0$
61	0 de 4	0	61	0	0	0	0
20	1 de 4	0,25	6,33	8,44	4,22	0,94	0,08
14	2 de 4	0,5	0,88	3,50	5,25	3,50	0,88
12	3 de 4	0,75	0,05	0,56	2,53	5,06	3,80
13	4 de 4	1	0	0	0	0	13
			68,25	12,5	12	9,5	17,75

Tabla 3.5

Se observa que la distribución de contactos final no se corresponde con los datos originarios (en negrilla ambos). Por ejemplo, habiendo 14 millones con probabilidad 0'5 (pues reciben 2 de 4 exposiciones) el modelo estima que 12 millones reciben dos contactos. Se incumple el criterio de reproducir fielmente los datos de partida. En concreto subestima el alcance total y toda la distribución de contactos excepto en el último, que está sobrestimado. Por otra parte, cumple los otros criterios, de solución única y coherencia de resultados. Greene critica los programas de planificación que aplican este método, ya que estarían produciendo errores sistemáticos.

2. Método de la probabilidad fija

Las probabilidades de exposición se determinan, igual que en el modelo actuarial, por el cociente entre el número de exposiciones recibidas y las exposiciones posibles de un periodo. La diferencia radica en que la distribución binomial de cada grupo se va a expandir con un F_i distinto del número de componentes del grupo. Con el ejemplo anterior, Schreiber sigue un proceso inverso: conocida la distribución de contactos total (E_k^n) y con probabilidades fijas (0, 0'25, 0'5, 0'75 y 1), se calculan los F_i individuos a los que asignar cada probabilidad, a través de *ecuaciones simultáneas*:

$$\begin{aligned} 61 &= F_1 + F_2 \cdot (0,75)^4 + F_3 \cdot (0,5)^4 + F_4 \cdot (0,25)^4 \\ 20 &= 4 \cdot F_2 \cdot (0,25) \cdot (0,75)^3 + 4 \cdot F_3 \cdot (0,5)^4 + 4 \cdot F_4 \cdot (0,25)^3 \cdot (0,75) \\ 14 &= 4 \cdot F_2 \cdot (0,25)^2 \cdot (0,75)^2 + 6 \cdot F_3 \cdot (0,5)^4 + 6 \cdot F_4 \cdot (0,25)^2 \cdot (0,75)^2 \\ 12 &= 4 \cdot F_2 \cdot (0,25)^3 \cdot (0,75) + 4 \cdot F_3 \cdot (0,5)^4 + 4 \cdot F_4 \cdot (0,75)^3 \cdot (0,25) \\ 13 &= F_2 \cdot (0,25)^4 + F_3 \cdot (0,5)^4 + F_4 \cdot (0,75)^4 + F_5 \end{aligned}$$

Resolviendo ese sistema de 5 ecuaciones con 5 incógnitas, se obtiene el número de personas que tiene asignada cada probabilidad. Los resultados son: $F_1=46'4$, $F_2=46'2$, $F_3=-2'7$, $F_4=24,9$ y $F_5=5,1$ (todos en millones). Evidentemente es imposible asignar la tercera probabilidad a un número negativo de personas. No se está cumpliendo el criterio de coherencia de resultados, y por tanto se debe descartar el método, aunque reproduzca los datos originarios y tenga solución única.

3. Método empírico

Se basa en que los individuos con igual hábito de exposición tienen la misma probabilidad de exposición, pero dicha probabilidad es desconocida y no puede calcularse por el método actuarial. Se pueden plantear las anteriores ecuaciones simultáneas, pero da lugar al doble de incógnitas que de ecuaciones: ello genera infinitas soluciones, aunque sólo tienen sentido las que tengan probabilidades e individuos no negativos. Como dice Greene¹³², "el método se llama así porque empíricamente se puede tomar cualquier probabilidad". Se incumple ahora el criterio de unicidad de la solución. Dos investigadores podrían alcanzar distintas soluciones que satisfacen el método, pero ¿cuál es la que realmente explica el comportamiento?; en el periodo original reproducen igual las exposiciones, pero en periodos más amplios ¿cuál escoger si se generan distintas exposiciones? Ello conduce a que este método también deba ser desaconsejado.

Del anterior ejemplo, Schreiber recoge dos soluciones. Las series de probabilidades (0; 0'04; 0'35; 0'70; 0'99) y (0'01; 0'20; 0'35; 0'80; 0'88) reproducen igual los datos originales. Friedman¹³³ emplea la primera de esas series, y añade que los valores de esas probabilidades deben estar distantes (cfr. **IV Parte, § C.4**).

¹³² GREENE, Jerome D.: "Personal Media Probabilities". *Journal of Advertising Research*. Vol. 10. Octubre 1970. Págs. 13-4.

¹³³ FRIEDMAN, Lawrence: "Constructing a Media Simulation Model". *Journal of Advertising Research*. Vol. 10. Nº 4. Agosto 1970. Págs. 33-9.

4. Método de la probabilidad personal

Greene¹³⁴ presentó este modelo, cuya lógica "es preasignar las probabilidades personales a cada entrevistado en cada medio, y calcular la distribuciones de frecuencias de la campaña directamente de esas probabilidades, procesando cada entrevistado uno a uno, y acabándolo completamente antes de tomar el siguiente entrevistado". Pretende evitar las largas tablas cruzadas de duplicaciones, triplicaciones, etc, con un modelo sencillo que estima las distribuciones de contactos individuo a individuo.

Se basa en la hipótesis de que las probabilidades de exposición individual siguen una distribución Beta entre la población. La distribución Beta es continua, permite reproducir los datos originarios, tiene solución única y no puede dar resultados absurdos. Además la distribución Beta se calcula únicamente a partir de la audiencia acumulada del primer y segundo número: en el ejemplo anterior, bastaría con conocer los hábitos de exposición de dos números. Chandon¹³⁵ lo expone en dos etapas: primero el ajuste de la distribución Beta a los datos originales, estimando los parámetros A y B; segundo, asignar a cada individuo una probabilidad de exposición condicionada por el número de veces que ese individuo ha sido expuesto. La asignación ha de hacerse de tal forma que la población total se distribuya exactamente según una Beta. Como la distribución Beta es el conjugado de la distribución Binomial, la probabilidad de que sea p la probabilidad de que el individuo i haya recibido k de n exposiciones, se expresa:

$$P(p/k, n) = \beta(p/k, n-k) = \frac{(n-1)!}{(k-1)! \cdot (n-k-1)!} p^{k-1} (1-p)^{n-k-1}$$

La probabilidad de que la probabilidad de exposición del individuo i esté contenida entre en el intervalo (a,b) dado que ha sido expuesto a k entre n exposiciones:

¹³⁴ GREENE, Jerome D.: "Personal Media Probabilities". *Journal of Advertising Research*. Vol. 10. Octubre 1970. Págs. 12-8.

¹³⁵ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 269-79.

$$P(a < p < b / k, n) = \frac{(n-1)!}{(k-1)! \cdot (n-k-1)!} \int_a^b p^{k-1} (1-p)^{n-k-1} \quad 0 \leq a; b \leq 1 \text{ reales}$$

Previamente los datos del correspondiente soporte se habían ajustado a una distribución Beta, en la primera etapa. Entonces, por la aplicación del teorema de Bayes, la probabilidad de que la probabilidad del individuo i sea p , conociendo que el individuo i ha recibido k de n exposiciones de un soporte con parámetros A y B , vale:

$$P(p / k, n, A, B) = \frac{\Gamma(A+B+n)}{\Gamma(k+A) \Gamma(n+B-k)} p^{k+A-1} (1-p)^{n+B-k-1}$$

La media de la distribución Beta de la anterior expresión, es la probabilidad de exposición esperada de todos los individuos que, como i , han recibido k entre n exposiciones. La probabilidad de que la probabilidad de exposición del individuo i esté comprendida dentro del intervalo (a, b) viene así determinada:

$$P(a < p < b / k, n, A, B) = \frac{\Gamma(A+B+n)}{\Gamma(k+A) \Gamma(n+B-k)} \int_a^b p^{k+A-1} (1-p)^{n+B-k-1}$$

Las probabilidades individuales de exposición se pueden agrupar en intervalos (por ejemplo, en 20 intervalos de cinco puntos porcentuales, de 0 a 1). Cada entrevistado va a tener asignada una probabilidad, propia de uno de estos intervalos, para evitar las correlaciones entre medios. Así el método de la probabilidad personal es análogo al de la Beta Binomial entre varios segmentos. Greene había encontrado correlaciones positivas entre varias revistas, planteando la mejora del modelo.

Chandon realizó un experimento del que concluía que este modelo tiene validez predictiva. Simuló el comportamiento ante un soporte de tres poblaciones de 1000 individuos; cada población seguía distintas distribuciones, pero ninguna Beta. Concluyó que el modelo de probabilidad personal era más preciso que el actuarial.

5. Método del último periodo

Así llamado porque, además del hábito de exposición, mide la audiencia del último periodo. Según Chandon, este método era utilizado en Inglaterra por el Target Group Index, en sus estudios periódicos. En España lo está usando ODEC, a partir de la información recogida en los cuestionarios del EGM. En función de la frecuencia y de la última exposición, se asigna una probabilidad a cada individuo. Dicho proceso, que suele conocerse como *probabilización*, primero agrupa las respuestas individuales, después calcula las probabilidades por grupos y finalmente asigna esas probabilidades a los componentes del grupo. Veámoslo a continuación, imaginando una muestra de 1.000 personas, a las que se ha preguntado:

1.- ¿cuántos números de ... (revista semanal) lee u ojea habitualmente, durante cuatro semanas? ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4

2.- ¿ha leído u ojeado el último número de ...? ☐ sí ☐ no

En la **Tabla 3.6**, los 1.000 individuos se agrupan según su hábito de exposición (primera respuesta). De cada grupo se recuentan los que han visto o leído el último número (segunda respuesta). Finalmente aparece la probabilidad de exposición, resultado de dividir la tercera columna entre la segunda; esa probabilidad se va a asignar a cada componente del grupo que ha declarado tener tal hábito de exposición¹³⁶.

¹³⁶ ZAMBRANO, LOURDES. "La probabilización en el Estudio General de Medios". Ponencia presentada en el 2º Seminario sobre Medios Impresos (54º Seminario de AEDEMO). Madrid. Diciembre 1990. Resumen de ponencias. Págs. 43-4.

Hábito de exposición	Nº de individuos	De ellos los expuestos la última vez	Probabilidad de exposición
0 de 4	350	5	0,01 (5/350)
1 de 4	125	30	0,24 (30/125)
2 de 4	145	70	0,48 (70/145)
3 de 4	150	110	0,73 (110/150)
4 de 4	230	220	0,96 (220/230)
TOTAL	1.000	435	

Tabla 3.6

El método del último periodo cumple los criterios de solución única y resultados lógicos, pero no reproduce correctamente los datos originales, como se ve simplemente calculando el alcance de los cuatro números:

$$C_4 = 350(1 - 0,99^4) + 125(1 - 0,76^4) + 145(1 - 0,52^4) + 150(1 - 0,27^4) + 230(1 - 0,04^4) \approx 617$$

Se ha subestimado el alcance observado en la encuesta, que era de 650 personas. También el método actuarial subestima el alcance, pero con un error de mayor cuantía:

$$C_4 = 350(1 - 1^4) + 125(1 - 0,75^4) + 145(1 - 0,5^4) + 150(1 - 0,25^4) + 230(1 - 0^4) \approx 601$$

El método del último periodo estima con mayor precisión el alcance acumulado y la distribución de contactos, en comparación con el método actuarial. Chandon además señala que le afecta menos la sobrevaloración que los entrevistados hacen de su frecuencia de exposición. Sin embargo, el método de probabilidad personal estima todavía mejor el alcance y la distribución de contactos, aunque la frecuencia sea sobrevalorada.

En la medición de los hábitos de audiencia mediante encuesta, se cometen diversos errores que, según Chandon, se originan de seis formas: la memoria de los entrevistados; la existencia de audiencia paralela y replicada; la explicación de un comportamiento continuo sobre la consulta de un periodo limitado; la agregación de probabilidades de exposición heterogéneas; las diferencias en la interpretación de las

posibles respuestas, especialmente si se emplean escalas verbales; y la inestabilidad debida a la no estacionariedad de las inserciones.

Los errores generados por las dos primeras pueden reducirse con el método del último periodo: el entrevistado manifiesta su hábito de exposición al medio, pero también su exposición en el último periodo, suavizando las sobre/infra-valoraciones del hábito. Martínez Ramos¹³⁷ describió tres de los problemas: los lectores paralelos, los lectores replicados y del efecto "telescopio" (efecto de la sobrevaloración de la audiencia); también planteaba algunas soluciones, como adecuar el método de panel *FTRY* (First Time Readership Yesterday) a las encuestas puntuales.

¹³⁷ MARTÍNEZ RAMOS, Emilio. "Problemas en la estimación de la audiencia de las revistas y diarios en una encuesta tradicional". Ponencia presentada en el 2º Seminario sobre Medios Impresos (54º Seminario de AEDEMO). Madrid. Diciembre 1990. Resumen de ponencias. Págs. 7-33.

PARTE IV

LA AUDIENCIA NETA ACUMULADA

INTRODUCCIÓN

La IV Parte se dedica a los modelos que miden el alcance y la distribución de contactos de campañas multi-soportes y multi-periodos. Encaramos el último escalón y, aunando las orientaciones anteriores, llegamos a las situaciones más comunes y, por tanto, más interesantes desde un punto de vista práctico. El estudio de la llamada *audiencia neta acumula* tiene dos enfoques, según que los algoritmos operen con datos globales o individuales: en el primer caso tendremos *modelos estocásticos* y en el segundo *modelos de simulación*. Su desarrollo histórico va parejo a la presentación de la información y a la evolución de los ordenadores: los estocásticos surgieron cuando los datos se publicaban de toda la población o de varios segmentos; después los ordenadores permitieron aislar la audiencia de cada individuo y componer específicamente los targets. Veremos que cada enfoque se puede llevar a cabo de diferentes formas.

Se recogen aquí diez modelos estocásticos, procurando que el desarrollo teórico esté acompañado de un ejemplo; no se aplica sobre los datos del EGM que, dado su carácter periódico, no mide la acumulación de audiencias. Sí se indican los resultados encontrados por otros autores; se observa una tendencia general: a mayor complejidad y a mayor necesidad de información, las estimaciones se hacen más precisas. También se describen seis modelos de simulación, a los que habría que añadir el TOM y el MARKSEL (I Parte, §C). Generalmente los modelos de simulación no se limitan a calcular el alcance y la distribución de contactos, sino que hacen planteamientos más amplios: buscar la campaña óptima, medir la evolución del comportamiento del consumidor, evaluar la efectividad, el olvido, etc; generalmente se vinculan a fuentes de información más extensas; esos modelos pueden perfeccionarse constantemente, pero aquí se exponen como aparecen en los artículos de referencia; en algunos casos no son modelos cerrados, sino que el planificador puede configurar algunas hipótesis del programa.

A) MODELOS ESTOCÁSTICOS

Conceptos previos e hipótesis implícitas

Se incluyen en este apartado diez modelos que calculan el alcance y la distribución de contactos a partir de datos agregados. Se llaman *estocásticos* porque, en todo o en parte del proceso, la audiencia se considera como un fenómeno probabilístico. Se basan en diversas hipótesis simplificadoras de la realidad, que hacen más sencillo el proceso pero, en contrapartida, producen estimaciones más alejadas de las observaciones reales. Chandon¹³⁸ indicaba las hipótesis implícitas de cada modelos (cfr. **Tabla 4.1**). La *estacionariedad*, o constancia de la probabilidad de exposición durante el tiempo, es aceptada por todos excepto por el modelo de Markov-Bernouilli. La *homogeneidad de la población* supone que los individuos tienen igual probabilidad de exposición. La *homogeneidad de los soportes* significa que todos tienen la misma probabilidad de audiencia, y es asumida por el Binomial, Markov-Bernouilli y Metheringham. Particularmente interesantes son las combinaciones de las hipótesis de *duplicación* (entre soportes) y de *acumulación* (intra soportes, entre varios periodos): pueden ser fenómenos aleatorios (si son independientes), no aleatorios o seguir alguna función de distribución. Las hipótesis realistas suelen requerir más información y algoritmos más complejos pero, en contrapartida, producen estimaciones más precisas.

Algunos tenían su correspondiente modelo de audiencia neta o de audiencia acumulada (Metheringham, Hofmans, etc) pero los demás, al reunir los dos enfoques, *constituyen el caso más general y pueden concretarse en los casos particulares*.

¹³⁸ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 440.

MODELO	Estaciona- riedad	Población	Soportes	Duplicación	Acumulación
Binomial	Estacionario	Homogénea	Homogéneos	Aleatoria	Aleatoria
Binomial Multivariable con independencia de soportes	Estacionario	Homogénea	Heterogéneos	Aleatoria	Aleatoria
Multinomial	Estacionario	Homogénea	Heterogéneos	No Aleatoria	Aleatoria
Markov-Bernouilli	No Estacionario	Homogénea	Homogéneos	No Aleatoria	No Aleatoria
Metherringham	Estacionario	Heterogénea	Homogéneos	Beta Media	Beta Media
Beta Matriz	Estacionario	Heterogénea	Heterogéneos	No Aleatoria	Beta Binomial
Multinomial Compuesto	Estacionario	Heterogénea	Heterogéneos	No Aleatoria	Beta Binomial
Hofmans	Estacionario	Heterogénea	Heterogéneos	No Aleatoria	No Aleatoria
Morgensztern	Estacionario	Heterogénea	Heterogéneos	No Aleatoria	No Aleatoria
Metherringham & Multinomial Compuesto	Estacionario	Heterogénea	Heterogéneos	No Aleatoria	No Aleatoria
Modelo Casual	Estacionario	Heterogénea	Heterogéneos	Beta Binomial	Beta Binomial

Tabla 4.1: Hipótesis de los Modelos Estocásticos. Fuente: Chandon (op. cit.).

1. Modelo Binomial

Como se comentaba en la segunda y tercera parte, éste es el modelo más sencillo pues se basa en las hipótesis menos restrictivas (véase la **Tabla 4.1**). Se aplica ahora *mutatis mutandis* a las combinaciones de múltiples soportes y múltiples inserciones. Seguimos recogiendo los datos de audiencia a nivel agregado. La hipótesis de estacionariedad hace referencia a que el fenómeno de la exposición permanece constante a lo largo del tiempo. La homogeneidad de la población supone que todos los individuos tienen la misma probabilidad de exposición a un soporte dado y, la homogeneidad de los soportes, que todos ellos tienen la misma probabilidad. La duplicación es aleatoria, luego se calcula como el producto de las dos audiencias implicadas, dividido por la población objetivo. Igualmente, la acumulación aleatoria hace que la probabilidad de contactar dos veces con un mismo soporte es el producto de contactar cada vez.

Dada la homogeneidad de la población y de los soportes, podemos referirnos a un soporte medio. La probabilidad de exposición a ese soporte medio se determina dividiendo la audiencia media entre la población objetivo. Esa probabilidad, por la hipótesis de estacionariedad, permanece constante en el tiempo. Por todo ello, cada nueva inserción de la campaña va a aportar lo mismo, independientemente del soporte implicado y del momento. Tampoco importan las inserciones ya acumuladas en ese soporte, porque toda inserción acumula lo mismo.

El alcance neto acumulado es el producto de la probabilidad de recibir al menos un contacto por la población objetivo:

$$C_n = P(x \geq 1 / p, n) \cdot P = [1 - P(s = 0 / p, n)] \cdot P$$

La distribución de contactos se define así, para k contactos entre n inserciones:

$$E_k^n = P(x = k / p, n) \cdot P = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \cdot P$$

Para llegar a estas expresiones se ha tenido que emplear la propiedad multiplicativa de la distribución binomial: cada soporte había generado por su cuenta una distribución binomial de contactos y, al tratarse de múltiples soportes, se genera otra distribución binomial, que supone el producto de todas las individuales.

Los modelos con múltiples inserciones y soportes fueron testados por Chandon para 57 campañas, a partir de encuestas de W. R. Simmons and Associates de 1972. Aunque Chandon considera que es una amplia serie de combinaciones, no comparto esa opinión. Además, como él mismo señala¹³⁹, tienen dos limitaciones: cada soporte no puede tener más de dos inserciones y las campañas sólo incluyen tres o cuatro soportes. Resultan demasiado restrictivas esas limitaciones. Para testar la estimación del alcance neto acumulado también emplea dos bases datos adicionales: una encuesta¹⁴⁰ de 5 revistas con un límite de 6 inserciones, y otra encuesta¹⁴¹ sobre 11 revistas.

La estimación del alcance neto acumulado, en las 57 campañas de Chandon, era muy deficiente: se obtenía una sobrestimación sistemática media del 31%, con desviaciones siempre superiores al intervalo de confianza. En la distribución de contactos se producía una sobrestimación del primer contacto mientras existían subestimaciones en los contactos más elevados. Hemos de tener en cuenta que estos errores aumentarían sustancialmente si las campañas tuvieran más soportes e inserciones: en ese caso, el efecto de hipótesis tan restrictivas acentúa la desviación de la estimación.

¹³⁹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 445-6.

¹⁴⁰ ALFRED POLITZ RESEARCH, Inc. "Life Study of Consumer Expenditures". 1959. New York.

¹⁴¹ ALFRED POLITZ MEDIA STUDIES. "Main Report, 1964 Magazine Study, the audiences of 11 magazines". Febrero de 1965. New York.

2. Modelo Binomial Multivariable con soportes independientes

2.1. Estructura y desarrollo

Modificando la hipótesis de homogeneidad de los soportes, el Modelo Binomial Multivariable con independencia de soportes mantiene el resto de hipótesis del Modelo Binomial. La heterogeneidad de los soportes, ahora supuesta, implica que la probabilidad de exposición es distinta entre soportes, de manera que ya no se puede hablar de soporte medio.

Se está suponiendo que cada soporte tiene unas probabilidades de exposición que siguen una variable de Bernouilli. Para múltiples soportes, las probabilidades de exposición conjunta se determinan por el producto de las distribuciones binomiales, debido a la hipótesis de independencia de soportes. Tratándose de m soportes (x_1, x_2, \dots, x_m), cada uno con sus correspondientes n_i inserciones, tendríamos las siguientes probabilidades de recibir k_i de contactos:

$$\begin{aligned}
 P(x_1 = k_1) &= \binom{n_1}{k_1} \cdot p_1^{k_1} \cdot q_1^{n_1 - k_1} & 0 \leq k_1 \leq n_1 \\
 P(x_2 = k_2) &= \binom{n_2}{k_2} \cdot p_2^{k_2} \cdot q_2^{n_2 - k_2} & 0 \leq k_2 \leq n_2 \\
 &\dots \\
 P(x_m = k_m) &= \binom{n_m}{k_m} \cdot p_m^{k_m} \cdot q_m^{n_m - k_m} & 0 \leq k_m \leq n_m
 \end{aligned}$$

La probabilidad conjunta de recibir esos contactos es:

$$\begin{aligned}
 P(x_1 = k_1; x_2 = k_2; \dots; x_m = k_m) &= P(x_1 = k_1) \cdot P(x_2 = k_2) \cdot \dots \cdot P(x_m = k_m) = \\
 &= \binom{n_1}{k_1} \cdot p_1^{k_1} \cdot q_1^{n_1 - k_1} \cdot \binom{n_2}{k_2} \cdot p_2^{k_2} \cdot q_2^{n_2 - k_2} \cdot \dots \cdot \binom{n_m}{k_m} \cdot p_m^{k_m} \cdot q_m^{n_m - k_m} = \prod_{i=1}^m \binom{n_i}{k_i} \cdot p_i^{k_i} \cdot q_i^{n_i - k_i}
 \end{aligned}$$

Pero lo que realmente nos interesa es la distribución de contactos final, es decir, los individuos que han recibido $1, 2, \dots, \sum_{i=1}^m n_i$ impactos, independientemente de si ha sido en el soporte $1, 2, \dots$, ó m . Para ello tendremos que sumar las probabilidades que tienen el mismo número de contactos. Sea s una nueva variable aleatoria que se define como la suma de x_1, x_2, \dots, x_m (no siendo ya binomial):

$$P(s=0) = P(x_1=0; x_2=0; \dots; x_m=0)$$

$$P(s=1) = P(x_1=1; x_2=0; \dots; x_m=0) + P(x_1=0; x_2=1; \dots; x_m=0) + \dots + P(x_1=0; x_2=0; \dots; x_m=1)$$

...

$$P(s=\sum_{i=1}^m n_i) = P(x_1=n_1; x_2=n_2; \dots; x_m=n_m)$$

El alcance neto acumulado se determina a partir de la probabilidad de recibir al menos un impacto o, dicho de otra forma, el complementario de la probabilidad de recibir cero impactos:

$$P(s \geq 1) = 1 - P(x_1=0; x_2=0; \dots; x_m=0) = 1 - \prod_{i=1}^m q_i^{n_i}$$

Las estimaciones del alcance neto acumulado y de la distribución de contactos resultaban tan deficientes como en el Modelo Binomial, para las 57 campañas testadas por Chandon (Simmons, 1972). Por ejemplo, el alcance se sobrestimaba en todos los casos, una media del 32%. Concluía que la hipótesis de heterogeneidad de los soportes no era tan relevante.

El Modelo Binomial Multivariable con independencia de soportes, en terminología de Chandon, corresponde al de Distribución Exacta según Lamas¹⁴², con la salvedad de que el primero usa los datos a nivel agregado. La Distribución Exacta se

¹⁴² LAMAS ALONSO, Carlos: "Eso de las evaluaciones no lo veo claro". *Investigación y Marketing*. Asociación Española de Estudios de Mercado, Marketing y Opinión (AEDEMO). Nº 42. Julio 1993. Pág. 60.

basa en el cálculo probabilístico individuo a individuo, y es considerado habitualmente en España como lo teóricamente más correcto. En la práctica no se usa pues, como señala Lamas, "los tiempos de respuesta de un ordenador se hacen inaceptables por lo que es necesario recurrir a simplificaciones del método teórico". Hoy en día, aun teniendo ordenadores tan potentes como los del centro de cálculo ODEC, no es asequible para planificar una campaña con numerosas inserciones en un tiempo razonable. A mi juicio, la empresa de los próximos años no es reducir ese tiempo de ordenador, sino dar un salto cualitativo, utilizar paneles de audiencia y no distribuir temporalmente las probabilidades.

2.2. Ejemplo

Para simplificar los cálculos supongamos un caso sencillo, aunque no sea una de las campañas testadas. Se incluyen dos inserciones en Muy Interesante y una en Teleprograma. Recordemos que tenían unas probabilidades de exposición de 0,213 y 0,1628 respectivamente. Vamos a denominar x_1 a Muy Interesante y x_2 a Teleprograma. Existen 6 combinaciones posibles de contactar con ese plan:

$$P(x_1 = 0; x_2 = 0) = (0,787)^2 \cdot (0,8372) \approx 0,5185$$

$$P(x_1 = 1; x_2 = 0) = 2 \cdot (0,787) \cdot (0,213) \cdot (0,8372) \approx 0,2807$$

$$P(x_1 = 0; x_2 = 1) = (0,787)^2 \cdot (0,1628) \approx 0,1008$$

$$P(x_1 = 1; x_2 = 1) = 2 \cdot (0,787) \cdot (0,213) \cdot (0,1628) \approx 0,0546$$

$$P(x_1 = 2; x_2 = 0) = (0,213)^2 \cdot (0,8372) \approx 0,038$$

$$P(x_1 = 2; x_2 = 1) = (0,213)^2 \cdot (0,1628) \approx 0,0074$$

Agregamos en la variable s , según el número de contactos recibidos:

$$P(s=0) = P(x_1 = 0; x_2 = 0) \approx 0,5185$$

$$P(s=1) = P(x_1 = 1; x_2 = 0) + P(x_1 = 0; x_2 = 1) \approx 0,2807 + 0,1008 \approx 0,3815$$

$$P(s=2) = P(x_1 = 1; x_2 = 1) + P(x_1 = 2; x_2 = 0) \approx 0,0546 + 0,038 \approx 0,0926$$

$$P(s=3) = P(x_1=2; x_2=1) \approx 0,0074$$

En porcentaje sobre el público objetivo, este modelo estima que el 51'85% recibe 0 contactos, el 38'15% 1 contacto, el 9'26% 2 contactos y el 0'74% 3 contactos. El alcance neto acumulado es la suma de los que al menos contactan una vez, es decir, el 48'15%.

2.3. Aplicación a nuestra investigación

De las 402 campañas referidas en la segunda parte, hemos extraído las 75 de cuatro soportes; las demás suponían cálculos excesivamente complejos. Aun siendo una muestra pequeña, se pueden extraer algunas conclusiones interesantes. Son campañas con una sola inserción por soporte de manera que, en nuestra terminología se les asocia la audiencia neta. Sin embargo ahora estamos más interesados en medir la distribución de contactos según este método. Hay que tener en cuenta que no se evalúa la hipótesis de acumulación libre, ya que los soportes sólo tienen una inserción.

En las **Tablas 4.2 y 4.3** recogemos las diferencias entre la distribución de contactos del TOM y la estimada por el Método Binomial Multivariable con independencia de soportes. La primera recoge separadamente cada uno de los contactos, mientras que la segunda busca alguna relación con el criterio de selección de soportes. No se puede analizar aquí la influencia de los públicos objetivos porque en cada segmento sólo teníamos cuatro casos.

Con- tactos	Media de la diferencia	Desv. típ. de la diferencia	Media de la diferencia absoluta	Desv. típ. de la diferencia absoluta	Media diferencia absoluta ponderada	Desv. típ. diferencia absoluta ponderada	Nº de casos
1	3,53%	1,84%	3,53%	1,84%	0,87%	0,53%	75
2	-0,23%	0,70%	0,62%	0,39%	0,03%	0,03%	75
3	-0,50%	0,48%	0,50%	0,47%	0,01%	0,01%	75
4	-0,08%	0,12%	0,08%	0,11%	0,00%	0,00%	75
TODOS	0,68%	0,35%	1,18%	0,62%	2,92%	1,41%	75

Tabla 4.2: Diferencias en la distribución de contactos del TOM y del Método Binomial Multivariable, según el número de contactos estimado.

El Modelo Binomial Multivariable sobrestima ligeramente la distribución de contactos del TOM, una media del 0'68% para todos los contactos (véase la última fila de la **Tabla 4.2**). En valor absoluto esa diferencia media llega a valer el 1'18%. Sin conceder la misma importancia a todos los contactos, la diferencia media absoluta ponderada roza el 3%, con una desviación típica de la mitad de ese valor: esos niveles empiezan a ser significativos. Pero son muy distintas las diferencias observadas entre los contactos. La mayor parte de los errores provienen del primer contacto (sobrestimación media del 3'53%), mientras que en los demás hay subestimaciones pequeñas y descendentes. Hay que señalar también que estas campañas agrupan los contactos principalmente en el primer nivel, porque constan de cuatro inserciones, una por soporte: la audiencia repetida *entre* soportes siempre es menor que la repetida *intra* soportes. Sucede precisamente que, ahí donde se agrupan los contactos, se producen las mayores diferencias.

Cri- terio	Media de la diferencia	Desv. tip. de la diferencia	Media de la diferencia absoluta	Desv. tip. de la diferencia absoluta	Media diferencia absoluta ponderada	Desv. tip. diferencia absoluta ponderada	Nº de casos
rentab	0,66%	0,38%	1,15%	0,64%	2,82%	1,52%	19
cobert	0,76%	0,36%	1,27%	0,67%	3,05%	1,48%	19
re+co	0,73%	0,40%	1,24%	0,72%	2,95%	1,61%	19
afin.	0,56%	0,24%	1,07%	0,43%	2,84%	1,04%	18
TODOS	0,68%	0,35%	1,18%	0,62%	2,92%	1,41%	75

Tabla 4.3: Diferencias en la distribución de contactos del TOM y del Método Binomial Multivariable, según el criterio de selección de soportes.

No existe relación entre el criterio de selección de soportes y las diferencias entre el TOM y el Método Binomial Multivariable (cfr. **Tabla 4.3**). Las campañas basadas en la rentabilidad, en la cobertura, etc, muestran los mismos niveles que todo el conjunto, ya sea teniendo en cuenta el signo, en valor absoluto, o ponderando el peso del contacto. Eso reflejan las medias de los correspondientes criterios, con 19 casos, uno por público objetivo, ó 18 casos en el criterio de afinidad, donde no tiene sentido el público objetivo general.

3. Modelo Multinomial

3.1. Fundamentos e hipótesis

El Modelo Multinomial se diferencia del Modelo Binomial Multivariable en la hipótesis de independencia de los soportes. En este caso, la duplicación no es libre de modo que la exposición a un soporte condiciona la exposición a otro. Mantiene las hipótesis de homogeneidad de la población, heterogeneidad de los soportes, acumulación aleatoria y estacionariedad de las inserciones. También sigue el tratamiento agregado de los individuos. En principio se limita a campañas con igual número de inserciones por soporte, aunque luego veremos que esto puede subsanarse.

Chandon describe que la Multinomial con múltiples variables requiere el conocimiento de las duplicaciones, triplicaciones, cuadruplicaciones,..., y los cálculos se hacen extremadamente complejos. Propone sustituirla por una Multinomial univariable, de la forma descrita por Steyn¹⁴³. Supongamos que en una semana se coloca una inserción en cada uno de los m soportes. Tendremos una probabilidad p_0 de no contactar con ningún soporte, una probabilidad p_1 de contactar con un soporte cualquiera, una probabilidad p_2 de contactar con dos soportes cualesquiera,..., hasta la probabilidad p_m de contactar con los m soportes. La probabilidad de que en n semanas de la campaña se produzcan exactamente K contactos sigue una distribución multinomial univariable:

$$P(K / n, p_1, p_2, \dots, p_m) = \sum \frac{n!}{k_0! \cdot k_1! \cdot k_2! \cdot \dots \cdot k_m!} p_0^{k_0} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_m^{k_m}$$

$$\forall K \in \{0, 1, 2, \dots, n \cdot m\}$$

$$\text{siendo } p_0 = 1 - p_1 - p_2 - \dots - p_m$$

¹⁴³ STEYN, H. S: "On the univariate series $F(t) = F(a, b_1, b_2, \dots, b_k, c, t, t^2, \dots, t^k)$ and its applications to probability theory". *Nederland Akademia Wetensch. Proceeding Series A*, 58. Págs. 588-95.

$$\text{y cumpliéndose que } \sum_{i=0}^m k_i = n \quad \text{y} \quad \sum_{i=1}^m i \cdot k_i = K$$

La probabilidad p_i de contactar exactamente con i soportes se calcula a través del ya comentado teorema de Waring.

$$p_i = S_i - \binom{i+1}{i} \cdot S_{i+1} + \binom{i+2}{i} \cdot S_{i+2} - \dots \pm \binom{m}{i} \cdot S_m$$

siendo S_0 igual a 0, y S_i igual a la suma de todas las probabilidades de exposición no exclusivas a i soportes.

Si todos los soportes no tienen el mismo número de inserciones, habría que atribuirles el número superior encontrado y, sobre esa hipótesis, emplear el Modelo Multinomial Univariable. Posteriormente se transformaría en la correcta distribución de contactos.

Cuando Chandon testó sus 57 campañas con este modelo, encontró cierta mejora con respecto a los dos anteriores, pero todavía con desviaciones considerables: se mantiene siempre la sobrestimación del alcance neto acumulado, que alcanza un valor medio del 25'2% y ninguna estimación se encuentra dentro del intervalo de confianza de la encuesta. En cuanto a la distribución de contactos, sistemáticamente sobrestimaba el primero y subestimaba los contactos más altos.

3.2. Ejemplo

Supongamos una campaña que conste de dos inserciones en cada uno de estos tres revistas: Muy Interesante, Teleprograma y El Jueves. En la **Tabla 4.4**, recordamos las audiencias y duplicaciones, sobre un *target* de 1.455.001 individuos.

	Audiencia revista		Audiencia duplicada
<i>Muy Interesante</i>	309.918	<i>Muy Inter. - TP</i>	55.208
<i>Teleprograma</i>	236.902	<i>Muy Inter. - El Jueves</i>	47.445
<i>El Jueves</i>	130.265	<i>TP - El Jueves</i>	23.935
Total	677.085	Total	126.588

Tabla 4.4

Por la fórmula de Hofmans (**II Parte, § B.4**) llegamos al alcance de estas tres revistas, suponiendo una inserción por cada una:

$$C = \frac{\left(\sum_{i=1}^n A_i \right)^2}{\sum_{i=1}^n A_i + \sum_{i=1, i < j}^n k_{ij} D_{ij}} = \frac{(677.085)^2}{677.085 + 140.189} = 560.943$$

Determinamos S_3 y, a través del teorema de Waring, la distribución de contactos de ese plan de 3 inserciones (E_i^3). Se expresan después como las p_i probabilidades de recibir i contactos.

$$S_3 = C - S_1 + S_2 = 560.943 - 677.085 + 126.588 = 10.446$$

$$E_0^3 = P - S_1 + S_2 - S_3 = 894.058 \quad p_0 = E_0^3 / P = 0,6145$$

$$E_1^3 = S_1 - 2 \cdot S_2 + 3 \cdot S_3 = 455.248 \quad p_1 = E_1^3 / P = 0,3129$$

$$E_2^3 = S_2 - 3 \cdot S_3 = 95.249 \quad p_2 = E_2^3 / P = 0,0655$$

$$E_3^3 = S_3 = 10.446 \quad p_3 = E_3^3 / P = 0,0072$$

Esas p_i probabilidades de exposición deben transformarse en las verdaderas probabilidades $P(K)$, considerando todas las semanas. En la **Tabla 4.5**, las cuatro primeras columnas muestran las 10 combinaciones posibles de los ($m=3$) soportes con las ($n=2$) inserciones; en la séptima columna aparecen los ($n \cdot m$) contactos; finalmente las probabilidades agrupadas por niveles y los contactos estimados.

k_0	k_1	k_2	k_3	[1] $\frac{2!}{k_0! \cdot k_1! \cdot k_2! \cdot k_3!}$	[2] $p_0^{k_0} \cdot \dots \cdot p_3^{k_3}$	Nº Contacto $K = \sum_{i=1}^3 i \cdot k_i$	Prob. contact. $P(K) = \Sigma[1] \cdot [2]$	Contactos estimados $P(K) \cdot P$
2	0	0	0	1	0,3776	0	0,3776	549.374
1	1	0	0	2	0,1923	1	0,3845	559.474
0	2	0	0	1	0,0979	2	0,1783	259.496
1	0	1	0	2	0,0402	2		
1	0	0	1	2	0,0044	3	0,0498	72.442
0	1	1	0	2	0,0205	3		
0	0	2	0	1	0,0043	4	0,0088	12.772
0	1	0	1	2	0,0022	4		
0	0	1	1	2	0,0005	5	0,0009	1.368
0	0	0	2	1	5,2E-05	6	5,2E-05	75

Tabla 4.5: Ejemplo de la distribución de contactos estimada según el Método Multinomial.

4. Modelo de Markov-Bernouilli

4.1. Hipótesis y desarrollo

Este modelo, introducido por Klotz¹⁴⁴, es el único de los modelos estocásticos que no considera estacionarias las audiencias. La exposición en un periodo depende únicamente de la exposición en el periodo anterior. La población y los soportes se consideran homogéneos y, por tanto, se hace referencia a un soporte medio. Por último, los fenómenos de duplicación y acumulación no son aleatorios.

Sea $X_n=0$ la no exposición y $X_n=1$ la exposición, a la n inserción. La hipótesis de independencia hace que la probabilidad de exposición a la n inserción no dependa de $n-1$ inserción:

$$P(X_n = 1 / X_{n-1} = 1) = P(X_n = 1 / X_{n-1} = 0) = p$$

$$P(X_n = 0 / X_{n-1} = 1) = P(X_n = 0 / X_{n-1} = 0) = 1 - p = q$$

Pero este modelo se basa en la dependencia, que hace mayores las duplicaciones y las acumulaciones en comparación con la hipótesis de independencia. Se considera que la probabilidad de exposición es mayor que p , si hubo exposición en la anterior inserción, y menor si no la hubo. La siguiente probabilidad λ mide el tamaño de la duplicación y/o de la acumulación, según se trate respectivamente de dos inserciones de distintos y/o del mismo soporte:

$$P(X_n = 1 / X_{n-1} = 1) = \lambda \quad p < \lambda < 1$$

La dependencia sólo afecta a las duplicaciones y acumulaciones de segundo orden: quedan fuera las triplicaciones, cuadruplicaciones,..., y las acumulaciones de orden superior a tres. Por otra parte, como se supone un soporte medio, λ es el cociente

¹⁴⁴ KLOTZ, J. "Statistical Inference in Bernoulli Trials with Dependence". *The Annals of Statistics*. 1973. Vol. 1. Págs. 373-9.

entre la probabilidad de dobles exposiciones (duplicaciones y/o acumulaciones) y la probabilidad de exposiciones:

$$\lambda = P(X_n = 1 / X_{n-1} = 1) = \frac{P(X_n = 1; X_{n-1} = 1)}{P(X_n = 1)}$$

siendo $P(X_n = 1)$ la probabilidad no condicional de exposición (p). Por otra parte, en una campaña de η inserciones y μ soportes, existen $\binom{\eta \cdot \mu}{2}$ dobles exposiciones; de ellas, $\binom{\eta}{2} \cdot \mu$ son acumulaciones de segundo orden y $\binom{\mu}{2} \cdot \eta^2$ duplicaciones de parejas de soportes. La probabilidad de dobles exposiciones es la suma de acumulaciones y duplicaciones entre el número de dobles inserciones. Operando se llega al valor de λ :

$$\lambda = \frac{(\sum \text{Acumulaciones} + \sum \text{Duplicaciones}) / \binom{\eta \cdot \mu}{2}}{\sum \text{Audiencias} / \mu} = \frac{2 \cdot (\sum \text{Acumulaciones} + \sum \text{Duplicaciones})}{\eta \cdot (\eta \cdot \mu - 1) \cdot \sum \text{Audiencias}}$$

Se puede hallar también la probabilidad complementaria: la no exposición condicionada a la exposición anterior:

$$P(X_n = 1 / X_{n-1} = 1) + P(X_n = 0 / X_{n-1} = 1) = 1 \quad \Leftrightarrow \\ P(X_n = 0 / X_{n-1} = 1) = 1 - \lambda$$

Ahora bien, como se quiere que la probabilidad de exposición del vehículo compuesto permanezca constante entre una inserción y la siguiente, se puede escribir:

$$P(X_n = 1 / X_{n-1} = 1) \cdot P(X_{n-1} = 1) + P(X_n = 1 / X_{n-1} = 0) \cdot P(X_{n-1} = 0) = P(X = 1) = p$$

$$P(X_n = 0 / X_{n-1} = 1) \cdot P(X_{n-1} = 1) + P(X_n = 0 / X_{n-1} = 0) \cdot P(X_{n-1} = 0) = P(X = 0) = q$$

Resolviendo esas ecuaciones, se obtiene finalmente:

$$P(X_n = 1 / X_{n-1} = 0) = (1 - \lambda) \cdot p / q$$

$$P(X_n = 0 / X_{n-1} = 0) = (1 - 2p + \lambda p) / q$$

La **distribución de contactos** conjunta se determina agregando las probabilidades de recibir el mismo número de contactos; esas probabilidades se determinan únicamente a partir de λ y p :

$$\begin{aligned} P(X_1 = x_1; X_2 = x_2; \dots; X_n = x_n) &= \\ &= P(X_1 = x_1) \cdot P(X_2 = x_2 / X_1 = x_1) \cdot \dots \cdot P(X_n = x_n / X_{n-1} = x_{n-1}) \\ &= \lambda^r \cdot (1 - \lambda)^{2(s-r)-t} \cdot (1 - 2p + \lambda p)^{n-1-2s+t+r} \cdot p^{s-r} \cdot q^{-(n-2-s+t)} \\ \text{con } x &= 0, 1; \quad i = 1, 2, \dots, n \text{ inserciones} \end{aligned}$$

$$r = \sum_{i=2}^n x_{i-1} \cdot x_i; \quad s = \sum_{i=1}^n x_i; \quad t = x_1 + x_n$$

Para los casos extremos de exposición a todas las inserciones y no exposición a ninguna, el cálculo se simplifica notablemente:

$$P(X_1 = 0; X_2 = 0; \dots; X_n = 0) = (1 - 2p + \lambda p)^{n-1} \cdot q^{2-n}$$

$$P(X_1 = 1; X_2 = 1; \dots; X_n = 1) = \lambda^{n-1} \cdot p$$

El **alcance neto acumulado** es la parte del público objetivo que no ha recibido cero contactos y su expresión:

$$C_n = [1 - P(X_1 = 0; X_2 = 0; \dots; X_n = 0)] \cdot P = [1 - (1 - 2p + \lambda p)^{n-1} \cdot q^{2-n}] \cdot P$$

La aplicación de Chandon afianzaba que los métodos basados en la homogeneidad de la población sobrestiman sistemáticamente el alcance. Pero este modelo, con una sobrestimación media del 22'18%, era el más preciso de los referidos hasta ahora, sin duda por la relajación de las hipótesis de duplicación y acumulación aleatorias. Igual que los anteriores, presenta sobrestimaciones del primer contacto y subestimaciones de los contactos superiores.

4.2. Ejemplo

Tomamos el mismo enunciado que en el caso anterior. Este modelo necesita también las audiencias acumuladas en dos inserciones: esos datos no están disponibles en el EGM porque no es un estudio continuo, sino que probabiliza las audiencias durante los meses que ocupa una ola (véase **I Parte, §B**). Supondremos que los datos de la **Tabla 4.4** corresponden a una semana determinada, y que las audiencias acumuladas después de dos semanas son: Muy Interesante, 452.872; Teleprograma, 301.273; y El Jueves, 219.734. La suma de acumulaciones vale por tanto:

$$\Sigma \text{Acumulaciones} = 452.872 + 301.273 + 219.734 - 677.085 = 296.794$$

Con tres soportes y dos inserciones, calculamos el parámetro λ y la probabilidad media p :

$$\lambda = \frac{2 \cdot (296.794 + 126.588)}{2 \cdot (2 \cdot 3 - 1) \cdot 677.085} \approx 0,1251 \quad p = \frac{677.085}{3 \cdot 1.455.001} \approx 0,1551$$

Podemos ya calcular el alcance neto acumulado de las seis inserciones:

$$C_n = [1 - (1 - 2p + \lambda p)^{n-1} \cdot q^{2-n}] \cdot P = [1 - (1 - 2 \cdot 0,1551 + 0,1251 \cdot 0,1551)^5 \cdot 0,8449^4] \cdot 1.455.001 \approx 942.831$$

La distribución de contactos se calcula a partir de p y λ , para cada una de los 2^6 comportamientos posibles ante 6 exposiciones. Se hallan en cada caso los parámetros r , s y t , en función de las alternativas de los fenómenos x_i . El proceso puede seguirse en la **Tabla 4.6**:

x_1	x_2	x_3	x_4	x_5	x_6	$r = \sum_{i=2}^n x_{i-1} \cdot x_i$	$s = \sum_{i=1}^n x_i$	$t = x_1 + x_n$	$P(x_1, \dots, x_n)$	$P(s)$	Contactos
0	0	0	0	0	0	0	0	0	0,3520	0,352	512.170
1	0	0	0	0	0	0	1	1	0,0674	0,4156	604.722
0	1	0	0	0	0	0	1	0	0,0702		
0	0	1	0	0	0	0	1	0	0,0702		
0	0	0	1	0	0	0	1	0	0,0702		
0	0	0	0	1	0	0	1	0	0,0702		
0	0	0	0	0	1	0	1	1	0,0674		
1	1	0	0	0	0	1	2	1	0,0100	0,187	272.101
1	0	1	0	0	0	0	2	1	0,0134		
1	0	0	1	0	0	0	2	1	0,0134		
1	0	0	0	1	0	0	2	1	0,0134		
1	0	0	0	0	1	0	2	2	0,0129		
0	1	1	0	0	0	1	2	0	0,0105		
0	1	0	1	0	0	0	2	0	0,0140		
0	1	0	0	1	0	0	2	0	0,0140		
0	1	0	0	0	1	0	2	1	0,0134		
0	0	1	1	0	0	1	2	0	0,0105		
0	0	1	0	1	0	0	2	0	0,0140		
0	0	1	0	0	1	0	2	1	0,0134		
0	0	0	1	1	0	1	2	0	0,0105		
0	0	0	1	0	1	0	2	1	0,0134		
0	0	0	0	1	1	1	2	1	0,0100		
1	1	1	0	0	0	2	3	1	0,0015	0,0406	59.143
1	1	0	1	0	0	1	3	1	0,0020		
1	1	0	0	1	0	1	3	1	0,0020		
1	1	0	0	0	1	1	3	2	0,0019		
1	0	1	1	0	0	1	3	1	0,0020		
1	0	1	0	1	0	0	3	1	0,0027		
1	0	1	0	0	1	0	3	2	0,0026		
1	0	0	1	1	0	1	3	1	0,0020		
1	0	0	1	0	1	0	3	2	0,0026		
1	0	0	0	1	1	1	3	2	0,0019		
0	1	1	1	0	0	2	3	0	0,0016		
0	1	1	0	1	0	1	3	0	0,0021		
0	1	1	0	0	1	1	3	1	0,0020		
0	1	0	1	1	0	1	3	0	0,0021		
0	1	0	1	0	1	0	3	1	0,0027		
0	1	0	0	1	1	1	3	1	0,0020		
0	0	1	1	1	0	2	3	0	0,0016		
0	0	1	1	0	1	1	3	1	0,0020		
0	0	1	0	1	1	1	3	1	0,0020		
0	0	0	1	1	1	2	3	1	0,0015		

Tabla 4.6 (I): Ejemplo del modelo de distribución de contactos de Markov-Bernoulli.

x_1	x_2	x_3	x_4	x_5	x_6	$r = \sum_{i=2}^n x_{i-1} \cdot x_i$	$s = \sum_{i=1}^n x_i$	$t = x_1 + x_n$	$P(x_1, \dots, x_n)$	$P(s)$	Contactos
1	1	1	1	0	0	3	4	1	0,0002	0,0045	6.513
1	1	1	0	1	0	2	4	1	0,0003		
1	1	1	0	0	1	2	4	2	0,0003		
1	1	0	1	1	0	2	4	1	0,0003		
1	1	0	1	0	1	1	4	2	0,0004		
1	1	0	0	1	1	2	4	2	0,0003		
1	0	1	1	1	0	2	4	1	0,0003		
1	0	1	1	0	1	1	4	2	0,0004		
1	0	1	0	1	1	1	4	2	0,0004		
1	0	0	1	1	1	2	4	2	0,0003		
0	1	1	1	1	0	3	4	0	0,0002		
0	1	1	1	0	1	2	4	1	0,0003		
0	1	1	0	1	1	2	4	1	0,0003		
0	1	0	1	1	1	2	4	1	0,0003		
0	0	1	1	1	1	3	4	1	0,0002		
1	1	1	1	1	0	4	5	1	3,3E-05	0,0002	345
1	1	1	1	0	1	3	5	2	4,3E-05		
1	1	1	0	1	1	3	5	2	4,3E-05		
1	1	0	1	1	1	3	5	2	4,3E-05		
1	0	1	1	1	1	3	5	2	4,3E-05		
0	1	1	1	1	1	4	5	1	3,3E-05		
1	1	1	1	1	1	5	6	2	4,7E-06	5E-06	7
SUMA TOTAL									1	1	1.455.001

Tabla 4.6 (II): Ejemplo del modelo de distribución de contactos de Markov-Bernouilli.

5. Modelo de Metherringham

5.1. Hipótesis y estructura

Metherringham¹⁴⁵ presentó un modelo para medir la **audiencia neta acumulada**, que es una prolongación del modelo de audiencia neta (**II Parte, §B.9**). En cuanto a la acumulación, señala que las distintas inserciones en una publicación alcanzan siempre al mismo grupo de personas. Lo explica sobre una campaña en cuatro revistas (notadas como letras del abecedario) con inserciones en distintos momentos (notados como números subíndices); por ejemplo, una campaña de nueve anuncios podría ser: A₁, B₂, B₄, C₁, C₄, C₆, D₁, D₃ y D₅; se basa en las siguientes hipótesis:

a) La cobertura de soporte (p) permanece constante en las distintas inserciones (las probabilidades de exposición son estacionarias de inserción a inserción):

$$p_{B_2} = p_{B_4} \quad ; \quad p_{C_1} = p_{C_4} = p_{C_6} \quad ; \quad p_{D_1} = p_{D_3} = p_{D_5}$$

b) La cobertura neta de dos soportes dados permanece también constante independientemente de las distintas inserciones:

$$p_{A_1B_2} = p_{A_1B_4} \quad ; \quad p_{B_2C_1} = p_{B_2C_4} = p_{B_2C_6} = p_{B_4C_1} = p_{B_4C_4} = p_{B_4C_6} \quad ; \quad \dots$$

c) La cobertura acumulada de dos inserciones en un mismo soporte permanece constante cualesquiera que sean las inserciones:

$$p_{C_1C_4} = p_{C_1C_6} = p_{C_4C_6} \quad ; \quad p_{D_1D_3} = p_{D_1D_5} = p_{D_3D_5}$$

Además asumía las hipótesis de heterogeneidad de los individuos y de homogeneidad de los soportes. La homogeneidad de los soportes no significa necesariamente que un individuo tenga la misma probabilidad de exposición ante cualquier soporte, sino que las probabilidades individuales de exposición deben estar

¹⁴⁵ METHERRINGHAM, Richard A. "Measuring the Net Cumulative Coverage of a Print Campaign". *Journal of Advertising Research*. Diciembre 1964. Vol. 4. Págs. 23-28.

distribuidas de tal forma que todos los vehículos terminen con la misma distribución de probabilidades.

Para calcular la cobertura, es necesario conocer todas las duplicaciones posibles: suponiendo que n_i es el número de inserciones del soporte i y n_j el de j , para m soportes

existirán $\sum_{i=1}^m \binom{n_i}{2}$ duplicaciones entre inserciones del mismo soporte, y $\sum_{i=1}^m \sum_{j=i+1}^{m-1} n_i \cdot n_j$

duplicaciones entre inserciones de distintos soportes. La suma de ambas es igual al número de duplicaciones posibles del total de inserciones de la campaña:

$$\binom{N}{2} = \sum_{i=1}^m \binom{n_i}{2} + \sum_{i=1}^m \sum_{j=i+1}^{m-1} n_i \cdot n_j$$

Veámoslo en el ejemplo propuesto por Metheringham, para una campaña de 9 anuncios en 4 revistas:

$$\begin{aligned} \binom{9}{2} &= \sum_{i=1}^4 \binom{n_i}{2} + \sum_{i=1}^4 \sum_{j=i+1}^3 n_i \cdot n_j = \\ &= \binom{n_1}{2} + \binom{n_2}{2} + \binom{n_3}{2} + \binom{n_4}{2} + n_1 \cdot n_2 + n_1 \cdot n_3 + n_1 \cdot n_4 + n_2 \cdot n_3 + n_2 \cdot n_4 + n_3 \cdot n_4 = \\ &= \binom{1}{2} + \binom{2}{2} + \binom{3}{2} + \binom{3}{2} + 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 3 + 2 \cdot 3 + 2 \cdot 3 + 3 \cdot 3 = 36 \end{aligned}$$

Como recoge Metheringham, Hyett¹⁴⁶ mostró que la no cobertura acumulada de n inserciones de un soporte se aproxima fielmente al momento de orden n respecto al origen de una variable Beta. Asume una probabilidad personal (q) de no contactar con ninguna inserción del soporte, y además esas probabilidades están distribuidas en la población de acuerdo con una variable Beta, así definida:

$$\phi(q) = \frac{q^{L-1} \cdot (1-q)^{m-1}}{B(L, m)} ; \quad 0 < q < 1$$

¹⁴⁶ HYETT, G. P.: Paper read to the Statistics Seminar. London School of Economics. Febrero de 1958.

Sobre estas presunciones, la no cobertura acumulada q_n viene determinada:

$$q_n = \int_0^1 \phi(q) q^n dq$$

La integral converge si $L > 0$, $m > 0$. Además q_n es el momento de orden n respecto al origen de $\phi(q)$, y es igual a:

$$\frac{L \cdot (L+1) \cdot \dots \cdot (L+n-1)}{(L+m) \cdot (L+m+1) \cdot \dots \cdot (L+m+n-1)}$$

Nótese que corresponde a la fórmula de k_r , para $s=L$ y $t=L+m$. Los valores de los parámetros L y m son determinados a partir de q_1 (la no cobertura de la primera inserción) y q_2 (la no cobertura de dos inserciones). La secuencia de las no coberturas para n soportes, en el caso de n inserciones en el mismo soporte, queda transformada así (con la notación de Hyett):

$$k_1 = \frac{\sum q_i}{\binom{n}{1}} = \frac{\sum q_i}{n} = q_1$$

$$k_2 = \frac{\sum q_{ij}}{\binom{n}{2}} = q_2$$

...

$$k_r = \frac{\sum q_{ijk} \text{ (r inserciones)}}{\binom{n}{r}} = q_r$$

...

$$k_n = \frac{\sum q_{ijk \dots} \text{ (n inserciones)}}{\binom{n}{n}} = \sum q_{ijk} = q_n$$

Se está asumiendo que la audiencia de cualquier inserción es constante, lo mismo que sucede con dos inserciones, tres inserciones, etc.

La fórmula de Hyett fue testada en Estados Unidos y en Inglaterra. Las estimaciones resultaron precisas en todos los casos.

Existe una importante diferencia entre la fórmula de k_i de Metheringham y la q_i de Hyett. En la fórmula de Hyett los parámetros tienen que ser positivos, para que la integral converja: ello se cumple siempre *intra* soporte, pero *entre* soportes frecuentemente se obtienen valores de s y t negativos. Por lo tanto, la fórmula no se puede basar en una variable Beta ni en suposiciones acerca de la distribución de probabilidades. La fórmula de k_i es empírica, pudiendo ser sus parámetros positivos y negativos; en el caso especial de la duplicación *intra* soporte, coincide con q_i .

Schreiber¹⁴⁷ probó el modelo de audiencia neta acumulada de Metheringham, comparando sus estimaciones con los datos de Simmons¹⁴⁸, para seis campañas que constaban cada una de dos inserciones en una revista y dos en otra. En todos los casos se producían sobrestimaciones que rondaban entre el 1% y el 3%, aunque hay que tener en cuenta la limitación de la investigación. Señala que probablemente los errores aumenten al considerar más soportes. También afirma que la fórmula tiende a subestimar la duplicación, porque asume que la duplicación *inter* soportes aumenta, con el número de inserciones, aproximadamente igual que la duplicación *intra* soportes. Además, si la campaña consta de trece inserciones en sólo dos soportes, se espera menos cobertura que la estimada por Metheringham.

Cuando Chandon testó el Modelo de Metheringham¹⁴⁹, que también llamaba Modelo Multivariante Beta Binomial, encontró un error medio del 5,2%, sobrestimando el alcance en los 57 casos. Se obtuvieron 34 estimaciones dentro del intervalo de confianza del 95% de la muestra. Supone una indudable mejora en relación a los modelos anteriores, que se basaban en la homogeneidad de la población.

¹⁴⁷ SCHREIBER, Robert J.: "The Metheringham Method for Media Mix: An Evaluation". *Journal of Advertising Research*. Vol. 9. Nº 2. Junio 1969. Págs. 54-6.

¹⁴⁸ SIMMONS, W. R. and ASSOCIATES RESEARCH, INC. "Mass Markets and the Media Reaching Them- 1966 Magazine Reach and Frequency Report". New York. W. R. Simmons. 1966.

¹⁴⁹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 491-2.

En la **Parte I, §D**, hemos testado el modelo de Metheringham en 400 campañas multi-soporte y multi-periodo, comparándolo con los resultados aportados por los modelos de simulación TOM y MARKSEL. Ahora únicamente añadiré un ejemplo.

5.2. Ejemplo

Supongamos una campaña con 3 inserciones en Muy Interesante, 2 en Teleprograma y 1 en El Jueves. Las duplicaciones *entre*-soportes e *intra*-soportes, las mismas que antes, se resumen a continuación (en la diagonal están las duplicaciones del soporte consigo mismo, es decir, la audiencia común de dos periodos):

	Muy Interesante	Teleprograma	El Jueves
Muy Interesante	166.964	55.208	47.445
Teleprograma		172.531	23.935
El Jueves			40.796

Calculamos la audiencia y la duplicación de un hipotético soporte medio en esta campaña:

$$\bar{A} = \frac{3 \cdot 309.918 + 2 \cdot 236.902 + 1 \cdot 130.265}{6} \approx 255.637$$

$$\bar{D} = \frac{3 \cdot 166.964 + 1 \cdot 172.531 + 0 \cdot 40.796 + 6 \cdot 55.208 + 3 \cdot 47.445 + 2 \cdot 23.935}{15} \approx 79.658$$

Como el alcance de la segunda inserción en este soporte medio es 431.616 ($= 2\bar{A} - \bar{D}$), podemos estimar los parámetros A y B de la distribución:

$$A = \frac{255.637 \cdot (431.616 - 255.637)}{1.455.001 \cdot (2 \cdot 255.637 - 431.616) - 255.637^2} \approx 0'9$$

$$B = \frac{0'9 \cdot (1.455.001 - 255.637)}{255.637} \approx 4'2$$

La distribución de contactos resultante es la siguiente:

$$E_0^7 = \binom{7}{0} \cdot \frac{10'2 \cdot 9'2 \cdot 8'2 \cdot 7'2 \cdot 6'2 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 601.596$$

$$E_1^7 = \binom{7}{1} \cdot \frac{0'9 \cdot 9'2 \cdot 8'2 \cdot 7'2 \cdot 6'2 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 368.302$$

$$E_2^7 = \binom{7}{2} \cdot \frac{1'9 \cdot 0'9 \cdot 8'2 \cdot 7'2 \cdot 6'2 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 227.590$$

$$E_3^7 = \binom{7}{3} \cdot \frac{2'9 \cdot 1'9 \cdot 0'9 \cdot 7'2 \cdot 6'2 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 134.088$$

$$E_4^7 = \binom{7}{4} \cdot \frac{3'9 \cdot 2'9 \cdot 1'9 \cdot 0'9 \cdot 6'2 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 72.694$$

$$E_5^7 = \binom{7}{5} \cdot \frac{4'9 \cdot 3'9 \cdot 2'9 \cdot 1'9 \cdot 0'9 \cdot 5'2 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 34.539$$

$$E_6^7 = \binom{7}{6} \cdot \frac{5'9 \cdot 4'9 \cdot 3'9 \cdot 2'9 \cdot 1'9 \cdot 0'9 \cdot 4'2}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 13.103$$

$$E_7^7 = \binom{7}{7} \cdot \frac{6'9 \cdot 5'9 \cdot 4'9 \cdot 3'9 \cdot 2'9 \cdot 1'9 \cdot 0'9}{11'1 \cdot 10'1 \cdot 9'1 \cdot 8'1 \cdot 7'1 \cdot 6'1 \cdot 5'1} \cdot 1.455.001 \approx 3.089$$

6. Modelo Matriz Beta

El Modelo Matriz Beta, introducido por Greene y Stock¹⁵⁰, es otro de los que acepta la heterogeneidad de la población pero, a diferencia de Metheringham, también la heterogeneidad en los soportes (ya no tiene sentido hablar de audiencias o duplicaciones de un soporte medio). La duplicación no es aleatoria y la acumulación de cada soporte sigue su propia distribución beta binomial. Las probabilidades de exposición se mantienen estacionarias a lo largo del tiempo.

Explicaremos el modelo a partir de un ejemplo. Se trata de determinar la distribución de contactos de una campaña con 4 inserciones en Muy Interesante y 3 en Teleprograma, para dirigida al *target* ya citado. En el siguiente cuadro tenemos la información inicial necesaria, con las acumulaciones y duplicaciones cruzadas de dos inserciones en cada revista:

		Contactos de Teleprograma			
		0	1	2	Total
Contactos de Muy Interesante	0	794.763	108.336	99.030	1.002.129
	1	225.053	41.023	19.832	285.908
	2	133.912	23.172	9.880	166.964
	Total	1.153.728	172.531	128.742	1.455.001

Cada fila y cada columna de la matriz se distribuye, por hipótesis, según una distribución Beta. Así podemos estimar los parámetros a y b de las distribuciones por columnas. Primero se calcula el alcance acumulado de una inserción (R_1) y dos inserciones (R_2), mediante estas fórmulas (tomando como ejemplo la primera columna):

$$R_1 = (E_1^1 + 2 \cdot E_2^2) / 2 = (225.053 + 2 \cdot 133.912) / 2 = 246.439$$

$$R_2 = E_1^1 + E_2^2 = 225.053 + 133.912 = 358.965$$

¹⁵⁰ GREENE, J. D. and STOCK, J. S. "Advertising Reach and Frequency in Magazines". Marketmath Inc. and Reader's Digest Association. New York. 1967.

$$\hat{a} = \frac{R_1 \cdot (R_2 - R_1)}{P \cdot (2R_1 - R_2) - R_1^2} = \frac{246.439 \cdot (358.965 - 246.439)}{1.153.728 \cdot (2 \cdot 246.439 - 358.965) - 246.439^2} \approx 0,2957$$

$$\hat{b} = \hat{a} \cdot (P - R_1) / R_1 = 0,2957 \cdot (1.153.728 - 246.439) / 246.439 \approx 1,4504$$

Los demás resultados de Teleprograma se recogen en este cuadro:

Contactos de Teleprograma				
	0	1	2	Total
R_1	246.439	43.684	19.796	309.918
R_2	358.965	64.195	29.712	452.872
a	0,2957	0,4288	0,2230	0,3016
b	1,4504	13,8532	16,1705	1,1144

Determinados los parámetros a y b , ya conocemos cómo se acumula la audiencia de Muy Interesante y podemos proyectar la distribución de contactos hasta la inserción que nos interese; en este caso, se hacen 4 inserciones en Muy Interesante. Como antes, se muestra la fórmula y cómo se aplica en la primera columna:

$$E_k^n = \binom{n}{k} \cdot \frac{(a+k-1) \cdot (a+k-2) \cdot \dots \cdot a \cdot (b+n-k-1) \cdot (b+n-k-2) \cdot \dots \cdot b}{(a+b+n-1) \cdot \dots \cdot (a+b)} \cdot P$$

$$E_0^4 = \binom{4}{0} \cdot \frac{4,4504 \cdot 3,4504 \cdot 2,4504 \cdot 1,4504}{4,7461 \cdot 3,7461 \cdot 2,7461 \cdot 1,7461} \cdot 1.153.728 \approx 738.524$$

...

$$E_4^4 = \binom{4}{4} \cdot \frac{3,2957 \cdot 2,2957 \cdot 1,2957 \cdot 0,2957}{4,7461 \cdot 3,7461 \cdot 2,7461 \cdot 1,7461} \cdot 1.153.728 \approx 39.238$$

Igualmente se obtiene para todas las columnas:

		Contactos de Teleprograma			
		0	1	2	Total
Contactos de Muy Interesante	0	738.524	154.443	122.418	851.241
	1	196.312	15.718	5.697	249.615
	2	110.584	2.125	575	156.483
	3	69.071	232	50	113.557
	4	39.238	14	2	84.105
	Total	1.153.728	172.531	128.742	1.455.001

Como las filas también siguen una distribución Beta, análogamente se hallan los parámetros a y b para cada contacto de Muy Interesante:

	R_1	R_2	a	b
0	199.639	276.860	0,23957	0,78192
1	13.556	21.415	0,08603	1,49811
2	1.638	2.700	0,01992	1,88384
3	165	281	0,00342	2,34079
4	10	17	0,00033	2,89783
Total	215.008	301.273	0,13146	0,75815

Análogamente también, se extiende la distribución Beta hasta el número de inserciones requeridas de Teleprograma:

Contactos de Teleprograma					
	0	1	2	3	Total
0	528.840	136.624	95.040	90.737	851.241
1	222.723	16.433	7.144	3.316	249.615
2	152.998	2.355	833	298	156.483
3	113.187	267	80	23	113.557
4	84.082	17	4	1	84.105
Total	1.101.830	155.695	103.101	94.375	1.455.001

La distribución de contactos de la campaña en su conjunto será la suma de las diagonales (perpendiculares a la principal) de esa matriz:

Nº Contactos	Contactados
0	528.840
1	359.346
2	264.471
3	213.423
4	88.498
5	395
6	27
7	1
Total	1.455.001

Si se hubiera extendido Teleprograma antes que Muy Interesante (primero por columnas y después por filas) se obtendría una distribución de contactos ligeramente distinta, porque la estimación de la Beta no es nunca perfecta. Las diferencias son

mayores cuando hay más soportes. Por otra parte, los cálculos crecen exponencialmente, haciéndose inasequibles desde un punto de vista práctico. Este método ha resultado muy preciso en sus estimaciones del alcance y de la distribución de contactos, según algunas verificaciones de Chandon. Mejoraba el modelo de Metheringham, sin sobrestimar el alcance, y detectaba mejor los picos de la distribución.

El mayor inconveniente del Modelo Matriz Beta es que requiere una información inicial compleja y no disponible. Por ejemplo, una campaña sobre tres soportes requiere una matriz de exposiciones con 3 dimensiones y 27 celdas (3^m ; $m=n^\circ$ de soportes); evidentemente esa información resultaría tan costosa, que hace desaconsejable este modelo. Es inimaginable que esa información se tuviera que recoger a través de encuestas, para los soportes que actualmente se incluyen en el EGM. Se concluye pues que este modelo realiza estimaciones muy precisas, pero es inasequible en la práctica.

7. Modelo Multinomial Compuesto

7.1. Estructura básica

El Modelo Multinomial Compuesto es una extensión del Modelo de Metheringham, con la diferencia de que los soportes se consideran heterogéneos, distribuyéndose cada uno según su propia distribución Beta. Las hipótesis coinciden con las de Modelo Matriz Beta, pero necesita mucha menos información inicial. También se le ha llamado Modelo Multinomial Dirichlet o Modelo Multivariable Beta Binomial con heterogeneidad de soportes.

Para estudiar su estructura, supongamos la matriz de exposiciones de dos soportes:

		Soporte 1		
		No exposición	Exposición	Total
Soporte 2	No Exposición	p_0	p_1	$p_0 + p_1$
	Exposición	p_2	p_3	$p_2 + p_3$
	Total	$p_0 + p_2$	$p_1 + p_3$	$p_0 + p_1 + p_2 + p_3 = 1$

Las probabilidades p_0 , p_1 , p_2 y p_3 son mutuamente excluyentes y suman la unidad, denominándose *eventos exclusivos*¹⁵¹. Mosimann¹⁵² señala que siguen una distribución de Dirichlet, también conocida como la β -distribución básica 4-dimensional de Mauldon¹⁵³:

¹⁵¹ GREENE, J. D. and STOCK, J. S. "Advertising Reach and Frequency in Magazines". Marketmath Inc. and Reader's Digest Association. New York. 1967.

¹⁵² MOSIMANN, J. E. "On the Compound Multinomial Distribution, the Multivariate β -Distribution, and Correlations among Proportions". *Biometrika*. 1962. Vol. 49. Págs. 65-82.

¹⁵³ MAULDON, J. G. "A Generalization of the Beta Distribution". *Annals of Mathematical Statistics*. Vol. 30. Págs. 509-20. 1959.

$$P(p_0, p_1, p_2, p_3) = \frac{\Gamma(A_0 + A_1 + A_2 + A_3)}{\Gamma(A_0) \cdot \Gamma(A_1) \cdot \Gamma(A_2) \cdot \Gamma(A_3)} \cdot p_0^{A_0-1} \cdot p_1^{A_1-1} \cdot p_2^{A_2-1} \cdot p_3^{A_3-1}$$

$$\text{con } 0 \leq p_i \leq 1; \quad \sum_{i=0}^3 p_i = 1; \quad A_0, A_1, A_2, A_3 > 0 \text{ reales.}$$

Dados los valores p_0, p_1, p_2 y p_3 , las probabilidades de k_0 no exposiciones a ambos soportes, k_1 exposiciones sólo al soporte 1, k_2 exposiciones sólo al soporte 2 y k_3 exposiciones a ambos soportes, con n inserciones por soporte, siguen esta distribución multinomial:

$$P(k_0, k_1, k_2, k_3) = \frac{n!}{k_0! \cdot k_1! \cdot k_2! \cdot k_3!} \cdot p_0^{k_0} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

$$\text{con } \sum_{i=0}^3 k_i = n; \quad k_i = 0, 1, 2, \dots, n.$$

Sea $K = \{k_0, k_1, k_2, k_3\}$ el vector de exposiciones, $P = \{p_0, p_1, p_2, p_3\}$ el vector de probabilidades y $A = \{A_0, A_1, A_2, A_3\}$ el vector de parámetros. La distribución multinomial compuesta se expresa:

$$P(K/A, n) = \int_0^1 \int_0^{1-p_0} \int_0^{1-p_0-p_1} P(K/P) \cdot P(P/A) dp_0 dp_1 dp_2$$

Finalmente se obtiene:

$$P(K/A, n) = \frac{n!}{k_0! \cdot k_1! \cdot k_2! \cdot k_3!} \cdot \frac{\Gamma(\sum_{i=0}^3 A_i) \cdot \prod_{i=0}^3 \Gamma(k_i + A_i)}{\left[\prod_{i=0}^3 \Gamma(A_i) \right] \cdot \left[\Gamma(n + \sum_{i=0}^3 A_i) \right]} \Rightarrow$$

$$P(K/A, n) = \frac{n! \cdot [A_0(A_0+1)\dots(A_0+k_0-1)] \cdot [A_1(A_1+1)\dots(A_1+k_1-1)] \cdot \dots \cdot [A_3(A_3+1)\dots(A_3+k_3-1)]}{k_0! \cdot k_1! \cdot k_2! \cdot k_3! \cdot (A_0 + A_1 + A_2 + A_3) \cdot (A_0 + A_1 + A_2 + A_3 + 1) \cdot \dots \cdot (A_0 + A_1 + A_2 + A_3 + n - 1)}$$

La distribución de contactos se obtiene agregando todos los componentes no negativos k_1 , k_2 y k_3 , tal que $k_1 + k_2 + 2k_3 = S$. Por tanto:

$$P(S) = \sum_s P(K / A, n), \text{ con } S = 0, 1, \dots, 2n.$$

El razonamiento se aplica análogamente para m soportes, como se puede ver en la expresión de su distribución multinomial compuesta:

$$P(K) = \frac{n! \cdot [A_0(A_0 + 1) \dots (A_0 + k_0 - 1)] \dots [A_1(A_1 + 1) \dots (A_1 + k_1 - 1)] \dots [A_m(A_m + 1) \dots (A_m + k_m - 1)]}{\left(\prod_{i=0}^m k_i!\right) \cdot \left(\sum_{i=0}^m A_i\right) \cdot \left(1 + \sum_{i=0}^m A_i\right) \dots \left(n - 1 + \sum_{i=0}^m A_i\right)}$$

La distribución multinomial compuesta tiene varias propiedades:

a) Cuando se trata sólo de un soporte ($m=1$) la expresión se reduce a una distribución beta binomial. De ahí que la multinomial compuesta se pueda considerar como una beta binomial multivariable.

b) La distribución conjunta de cualquier subgrupo de soportes y todas las distribuciones condicionales y marginales, son también multinomiales compuestas. Esto significa que la distribución condicional de exposiciones a un vehículo, dado un particular grupo de exposiciones en todos los otros vehículos, sigue una distribución beta binomial.

c) La distribución de cualquier suma de variables multinomiales compuestas es otra distribución multinomial compuesta. Esto procede del hecho de que una suma de variables multinomiales sigue una distribución multinomial y que una suma de variables de Dirichlet sigue una variable Dirichlet.

Para la estimación de los parámetros, Mosimann estableció los momentos de la multinomial compuesta:

$$E(k_i) = \frac{n \cdot A_i}{s} \quad \text{con } s = \sum_{i=0}^m A_i$$

$$E(k_i k_j) = \frac{n \cdot (n-1) \cdot A_i \cdot A_j}{s \cdot (s+1)}$$

$$\text{Var}(k_i) = \frac{n \cdot A_i \cdot (s - A_i) \cdot (n + s)}{s^2 \cdot (s+1)}$$

$$\text{Cov}(k_i k_j) = \frac{-n \cdot A_i \cdot A_j \cdot (n + s)}{s^2 \cdot (s+1)}$$

$$\rho_{ij} = - \sqrt{\frac{A_i \cdot A_j}{(s - A_i) \cdot (s - A_j)}}$$

Las varianzas y covarianzas son similares a las de una multinomial común, excepto para el factor c , que toma el valor $(n+s)/(s+1)$. De ahí se deducen las estimaciones de A_i de la multinomial compuesta:

$$\hat{A}_i = \frac{\bar{k}_i \cdot (n - \hat{c})}{n \cdot (\hat{c} - 1)} \quad i = 1, 2, \dots, m$$

Esta serie de ecuaciones puede resolverse siempre que c haya sido estimado primero. La estimación de c tiene fácil solución teórica, pero en la práctica exigiría unos datos no disponibles. Por ello Chandon proponía una simplificación que le daba buenos resultados. Acabamos de definir c en función de n y s ; siendo n conocido, nos queda por saber s , que es la suma de los A_i . Este autor constató en sus encuestas que las sumas de los parámetros α y β de cada vehículo eran muy parecidas. Entonces s podría ser estimado como la media ponderada de la suma de los parámetros de cada vehículo, con ponderación proporcional a las respectivas audiencias individuales.

Con esa hipótesis, Chandon aplicó el Modelo Multinomial Compuesto a sus 57 campañas. Los resultados mejoraron sustancialmente a otros métodos. Al testar el alcance neto acumulado, el error medio absoluto alcanzó un 2'14%, dándose 44 sobrestimaciones y 13 subestimaciones. Había tendencia a la sobrestimación, pero no sistemática como en otros modelos. Además, en 49 de los casos, el error era menor que

el intervalo de confianza de la muestra. Las frecuencias de contactos también eran más precisas, y el modelo era capaz de detectar distribuciones con varios picos; contrariamente al modelo de Metheringham, sobrestima los contactos pares y subestima los impares. Esto llevó a Chandon a proponer un nuevo modelo, que calculara el alcance y las frecuencias como la media de las estimaciones del Modelo de Metheringham y del Multinomial Compuesto: como cabría esperar, la distribución de contactos se estimaba mejor porque la media compensaba las diferencias opuestas; sin embargo, el alcance seguía tendiendo a la sobrestimación, porque ambos modelos sobrestiman el alcance; así, había 36 sobrestimaciones frente a 21 subestimaciones, con un error absoluto medio del 2'73%. Además 50 casos estaban dentro del intervalo de confianza de la encuesta. Todo ello indica que la media de los modelos es más precisa que cada modelo en particular.

Otra descripción del Modelo Multinomial Compuesto, junto con un ejemplo, fue realizada por León¹⁵⁴, que proponía hacer un programa de ordenador para aplicar fácilmente el Modelo, una vez que se cuente con los datos de audiencia adecuados.

7.2. Ejemplo

Supongamos una campaña que consta de dos inserciones en Muy Interesante y dos en Teleprograma. Los datos que necesitamos son el alcance de la primera y segunda inserción en cada revista, y el alcance neto de ambas (entre paréntesis, el porcentaje sobre el público objetivo):

Revistas	Alcance 1ª inserción	Alcance 2ª inserción	Alcance neto del par
Muy Interesante	309.918 (21,3%)	452.872 (31,13%)	491.612 (33,79%)
Teleprograma	236.902 (16,28%)	301.273 (20,71%)	

¹⁵⁴ LEÓN, José Luis. "El Modelo Multinomial Compuesto Dirichlet. Un modelo para calcular la distribución de exposiciones de un plan de medios más exactamente". *Investigación y Marketing*. AEDEMO. Nº 21. Julio de 1986. Págs. 27-30.

Se calculan las probabilidades de exposición a las dos revistas, sabiendo que son fenómenos mutuamente excluyentes:

- No exposición a ambas revistas: $p_0 = 1 - 0,3379 \approx 0,6621$
- Exposición exclusiva a Muy Interesante: $p_1 = 0,3379 - 0,1628 \approx 0,1751$
- Exposición exclusiva a Teleprograma: $p_2 = 0,3379 - 0,213 \approx 0,1249$
- Exposición a ambas revistas: $p_3 = 0,213 - 0,1751 \approx 0,0379$

A continuación se calculan los parámetros beta de los soportes:

a) Muy Interesante:

$$\alpha_1 = \frac{0,213 \cdot (0,3113 - 0,213)}{2 \cdot 0,213 - (0,213)^2 - 0,3113} \approx 0,3016$$

$$\beta_1 = \frac{0,3016 \cdot (1 - 0,213)}{0,213} \approx 1,1144$$

b) Teleprograma:

$$\alpha_2 = \frac{0,1628 \cdot (0,2071 - 0,1628)}{2 \cdot 0,1628 - (0,1628)^2 - 0,2071} \approx 0,0782$$

$$\beta_2 = \frac{0,0782 \cdot (1 - 0,1628)}{0,1628} \approx 0,4023$$

La estimación del parámetro s resulta de la media ponderada de las sumas $\alpha_i + \beta_i$ correspondientes a cada revista.

$$\hat{s} = \frac{(0,3016 + 1,1144) \cdot 0,213 + (0,0782 + 0,4023) \cdot 0,1628}{0,213 + 0,1628} \approx 1,0108$$

Los valores de \hat{A}_i se estiman multiplicando las p_i por el parámetro s anterior:

$$\hat{A}_i = p_i \cdot \hat{s} \quad \hat{A}_0 \approx 0,6692 ; \hat{A}_1 \approx 0,1769 ; \hat{A}_2 \approx 0,1262 ; \hat{A}_3 \approx 0,0384$$

En la **Tabla 4.7**, se realizan los cálculos para determinar las personas alcanzadas en cada alternativa:

k_0	k_1	k_2	k_3	$\frac{n!}{k_0! \cdot k_1! \cdot k_2! \cdot k_3!}$ [1]	$\frac{[A_0(A_0+1)\dots(A_0+k_0-1)] \dots [A_3(A_3+1)\dots(A_3+k_3-1)]}{(A_0+A_1+A_2+A_3) \cdot (A_0+A_1+A_2+A_3+1)}$ [2]	Probab. [1] · [2]	Personas alcanzadas
2	0	0	0	1	$\frac{0,6692 \cdot 1,6692}{1,0108 \cdot 2,0108}$	0,5497	799.764
0	2	0	0	1	$\frac{0,1769 \cdot 1,1769}{1,0108 \cdot 2,0108}$	0,1025	149.087
0	0	2	0	1	$\frac{0,1262 \cdot 1,1262}{1,0108 \cdot 2,0108}$	0,0699	101.766
0	0	0	2	1	$\frac{0,0384 \cdot 1,0384}{1,0108 \cdot 2,0108}$	0,0196	28.509
1	1	0	0	2	$\frac{0,6692 \cdot 0,1769}{1,0108 \cdot 2,0108}$	0,1165	169.552
1	0	1	0	2	$\frac{0,6692 \cdot 0,1262}{1,0108 \cdot 2,0108}$	0,0831	120.948
1	0	0	1	2	$\frac{0,6692 \cdot 0,0384}{1,0108 \cdot 2,0108}$	0,0253	36.750
0	1	1	0	2	$\frac{0,1769 \cdot 0,1262}{1,0108 \cdot 2,0108}$	0,0220	31.977
0	1	0	1	2	$\frac{0,1769 \cdot 0,0384}{1,0108 \cdot 2,0108}$	0,0067	9.716
0	0	1	1	2	$\frac{0,1262 \cdot 0,0384}{1,0108 \cdot 2,0108}$	0,0048	6.931
SUMA TOTAL						1	1.455.001

Tabla 4.7: Ejemplo de cálculo del Modelo Multinomial Compuesto.

Esos resultados pueden ordenarse en la matriz de exposiciones:

		Contactos de Muy Interesante			
		0	1	2	Total
Contactos de Teleprograma	0	799.764	169.552	149.087	1.118.403
	1	120.948	68.727	9.716	199.391
	2	101.766	6.931	28.509	137.206
	Total	1.022.478	245.210	187.313	1.455.001

Y agregando las diagonales, se llega a la distribución de contactos conjunta:

Nº Contactos	Contactados
0	799.764
1	290.500
2	319.581
3	16.647
4	28.509
Total	1.455.001

8. Modelo de Hofmans

8.1. Explicación integral del fenómeno

Hofmans presentó también un modelo ad-hoc para determinar el alcance neto acumulado de una combinación de soportes¹⁵⁵. Consiste en una ecuación, que integra las fórmulas del alcance neto y del alcance acumulado (cfr. **II** y **III Partes**). El alcance neto acumulado de m soportes, con un total de N inserciones, se determina así:

$$C_{N,m} = \frac{\left(\sum_{i=1}^m n_i \cdot A_i \right)^2}{\sum_{i=1}^m n_i \cdot A_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^m n_i \cdot n_j \cdot k_{ij} \cdot A_{ij} + \sum_{i=1}^m \frac{n_i \cdot (n_i - 1)^a}{2} \cdot k_i \cdot d_i}$$

siendo:

$\sum_{i=1}^m n_i \cdot A_i$: la audiencia bruta o suma de audiencias.

A_i : la audiencia del soporte i ; $i = \{1, 2, 3, \dots, m\}$.

n_i : número de inserciones del soporte i .

$N = \sum_{i=1}^m n_i$: el número de inserciones de la campaña (N) es la suma de las n_i

inserciones en los m soportes.

$\sum_{i=1}^{m-1} \sum_{j=i+1}^m n_i \cdot n_j \cdot k_{ij} \cdot A_{ij}$: la suma de las duplicaciones entre distintos soportes y

distintas inserciones. La duplicación entre dos soportes i y j debe multiplicarse por el número de inserciones de cada soporte, como se ve en la **Figura 4.1**:

¹⁵⁵ HOFMANS, PIERRE. "Measuring the Cumulative Net Coverage of Any Combination of Media". *Journal of Marketing Research*. Agosto de 1966. Vol. 3. Págs. 275-78.

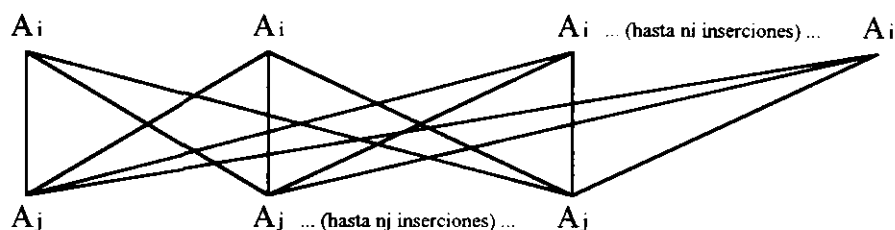


Figura 4.1

$$\sum_{i=1}^m \frac{n_i \cdot (n_i - 1)^a}{2} \cdot k_i \cdot d_i : \text{ la suma de las duplicaciones acumuladas para los}$$

distintos soportes.

Hofmans¹⁵⁶ presentó la estimación del alcance neto acumulado de dieciséis combinaciones de cinco revistas, a partir de los datos del estudio de Politz¹⁵⁷. Había colocado el mismo número de inserciones en cada soporte, para simplificar el cálculo. El error de estimación nunca superó el 1,7%, con un error medio del 0,9%, mostrándose así la acertada precisión del modelo. El autor señala que los cálculos se simplificarían bastante si se considera el valor de a igual para todas las revistas, por ejemplo su valor medio. Si eso no se admitía, era necesario conocer la audiencia acumulada de la tercera inserción.

Chandon señala que la fórmula tiene algunos inconvenientes: la parte de la *acumulación* tiende a la sobrestimación, y su cálculo requiere datos del alcance acumulado de las tres primeras inserciones; la parte de la *duplicación* también produce sobrestimaciones; como ambas partes están en el denominador, finalmente el alcance neto acumulado está subestimado; por otra parte, el modelo no calcula la distribución de contactos.

¹⁵⁶ HOFMANS, Pierre. "Une Méthode Pratique pour calculer la Couverture Nette Cumulée d'une Série de Supports". París. IREP, Study N° 12. Junio de 1965.

¹⁵⁷ ALFRED POLITZ RESEARCH, INC. "Life Study of Consumer Expenditures". New Time, Inc. 1959.

Chandon propuso una modificación en el cálculo de la acumulación, para no requerir los datos de alcance de la tercera inserción. Recordemos la fórmula del alcance acumulado de n inserciones en un soporte:

$$R_n = \frac{(n \cdot R_1)^2}{n \cdot R_1 + \frac{n \cdot (n-1)^a}{2} \cdot k \cdot d} \quad \text{con} \quad k = \frac{2 \cdot R_1}{R_2} \quad \text{y} \quad d = 2 \cdot R_1 - R_2$$

Despejando $k \cdot d$ de la primera fórmula, se consigue expresar en función del número de inserciones, del parámetro a , del alcance de la primera inserción y del alcance de las n inserciones:

$$k \cdot d = \frac{2 \cdot R_1 \cdot (n \cdot R_1 - R_n)}{R_n \cdot (n-1)^a}$$

Igualmente, despejando el parámetro a , y sustituyendo los valores de k y d , se obtiene la siguiente expresión:

$$a = \frac{\log \left[\frac{R_2 \cdot (n \cdot R_1 - R_n)}{R_n \cdot (2 \cdot R_1 - R_2)} \right]}{\log(n-1)}$$

De esta manera, el fenómeno de la acumulación se explica sin necesidad del alcance acumulado de la tercera inserción (R_3). Si es necesario conocer R_n , y Chandon propone hallarlo a partir de otro de los modelos de acumulación, como el de Agostini o el Beta Binomial, a partir únicamente de R_1 y R_2 (véase **III Parte**).

Chandon testó así el modelo de acumulación neta de Hofmans, para 57 campañas¹⁵⁸. El porcentaje de error medio absoluto era del 2,8%, produciendo sobrestimaciones en 22 casos y subestimaciones en 35 casos. El error estaba dentro del

¹⁵⁸ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 534-5.

intervalo de confianza de la encuesta en 47 ocasiones. Estimaba el alcance mejor que el modelo de Metheringham pero peor que el Multinomial Compuesto.

8.2. Ejemplo

Supongamos una campaña que conste de 5 inserciones en Muy Interesante, 4 en Teleprograma y 3 en El Jueves. Vamos a hallar el alcance neto acumulado a partir de los datos de la **Tabla 4.4** y del ejemplo de Markov-Bernoulli. Notamos R_1 y R_2 como el alcance acumulado de 1 y 2 inserciones, D_{ij} la duplicación con su correspondiente coeficiente k_{ij} :

	R_1	R_2		D_{ij}	k_{ij}
<i>Muy Interesante</i>	309.918	452.872	<i>Muy Int. - TP</i>	55.208	1,1123
<i>Teleprograma</i>	236.902	301.273	<i>Muy Int. - Jueves</i>	47.445	1,1208
<i>El Jueves</i>	130.265	219.734	<i>TP - Jueves</i>	23.935	1,0697

Hallamos el alcance acumulado total por revista, a partir del modelo de acumulación de Agostini:

a) Muy Interesante:

$$a = \frac{1}{\log 2} \cdot \log \frac{R_1 \cdot (1 - R_1)}{R_2 - R_1} = \frac{1}{\log 2} \cdot \log \frac{0,213 \cdot (1 - 0,213)}{0,3113 - 0,213} \approx 0,7708$$

$$R_5 = 1 - (1 - 0,213) \cdot \left(1 - \frac{0,213}{2^{0,7708}}\right) \cdot \left(1 - \frac{0,213}{3^{0,7708}}\right) \cdot \left(1 - \frac{0,213}{4^{0,7708}}\right) \cdot \left(1 - \frac{0,213}{5^{0,7708}}\right) \approx 0,4557$$

b) Teleprograma:

$$a = \frac{1}{\log 2} \cdot \log \frac{0,1628 \cdot (1 - 0,1628)}{0,2071 - 0,1628} \approx 1,6234$$

$$R_4 = 1 - (1 - 0,1628) \cdot \left(1 - \frac{0,1628}{2^{1,6234}}\right) \cdot \left(1 - \frac{0,1628}{3^{1,6234}}\right) \cdot \left(1 - \frac{0,1628}{4^{1,6234}}\right) \approx 0,242$$

c) El Jueves:

$$a = \frac{1}{\log 2} \cdot \log \frac{0,0895 \cdot (1 - 0,0895)}{0,1510 - 0,0895} \approx 0,4067$$

$$R_3 = 1 - (1 - 0,0895) \cdot \left(1 - \frac{0,0895}{2^{0,4067}}\right) \cdot \left(1 - \frac{0,0895}{3^{0,4967}}\right) \approx 0,1996$$

Ya podemos calcular acumulación de cada soporte según el modelo de Hofmans, y luego sumar su total:

a) Muy Interesante:

$$a_1 = \frac{\log \left[\frac{R_2 \cdot (n_1 \cdot R_1 - R_n)}{R_n \cdot (2 \cdot R_1 - R_2)} \right]}{\log(n-1)} = \frac{\log \left[\frac{0,3113 \cdot (5 \cdot 0,213 - 0,4557)}{0,4557 \cdot (2 \cdot 0,213 - 0,3113)} \right]}{\log 4} \approx 0,9294$$

$$k_1 \cdot d_1 = \frac{2 \cdot R_1 \cdot (n_1 \cdot R_1 - R_n)}{R_n \cdot (n_1 - 1)^a} = \frac{2 \cdot 0,213 \cdot (5 \cdot 0,213 - 0,4557)}{0,4557 \cdot 4^{0,9294}} \approx 0,1571$$

b) Teleprograma:

$$a_2 = \frac{\log \left[\frac{0,2071 \cdot (4 \cdot 0,1628 - 0,242)}{0,242 \cdot (2 \cdot 0,1628 - 0,2071)} \right]}{\log 3} \approx 0,9858$$

$$k_2 \cdot d_2 = \frac{2 \cdot 0,1628 \cdot (4 \cdot 0,1628 - 0,242)}{0,242 \cdot 3^{0,9858}} \approx 0,1865$$

c) El Jueves:

$$a_3 = \frac{\log \left[\frac{0,151 \cdot (3 \cdot 0,0895 - 0,1996)}{0,1996 \cdot (2 \cdot 0,0895 - 0,151)} \right]}{\log 2} \approx 0,8954$$

$$k_3 \cdot d_3 = \frac{2 \cdot 0,0895 \cdot (3 \cdot 0,0895 - 0,1996)}{0,1996 \cdot 2^{0,8954}} \approx 0,0332$$

$$\sum_{i=1}^3 \frac{n_i \cdot (n_i - 1)^{a_i}}{2} \cdot k_i \cdot d_i = \frac{5 \cdot 4^{0,9294}}{2} \cdot 0,1571 + \frac{4 \cdot 3^{0,9858}}{2} \cdot 0,1865 + \frac{3 \cdot 2^{0,8954}}{2} \cdot 0,0332 \approx 2,6184$$

Por otra parte, hemos determinado el alcance bruto y la suma de las duplicaciones *entre* soportes:

$$\sum_{i=1}^3 n_i \cdot A_i = 5 \cdot 309.918 + 4 \cdot 236.902 + 3 \cdot 130.265 = 3.018.258$$

$$\sum_{i=1}^2 \sum_{j=i+1}^3 n_i \cdot n_j \cdot k_{ij} \cdot A_{ij} = 5 \cdot 4 \cdot 1,1123 \cdot 55.208 + 5 \cdot 3 \cdot 1,1208 \cdot 47.445 + 4 \cdot 3 \cdot 1,0697 \cdot 23.935 \approx 2.333.042$$

Sustituyendo en la fórmula del alcance neto acumulado, se obtiene finalmente:

$$C_{12,3} = \frac{(3.018.258)^2}{3.018.258 + 2.330.042 + 2,6184 \cdot 1.455.001} \approx 994.412$$

9. Modelo de Morgensztern

9.1. Descripción del proceso

El Modelo de Morgensztern hace las mismas hipótesis que el de Hofmans, como refleja la **Tabla 4.1**: heterogeneidad de los individuos y de los soportes, duplicación y acumulación no aleatorias, y estacionariedad de las exposiciones entre inserciones. Su formulación es de hecho una modificación de la fórmula de Hofmans.

Nos referiremos primero a **dos soportes**. Morgensztern¹⁵⁹ calcula su alcance neto acumulado, basándose en esta hipótesis: la duplicación neta acumulada dividida por la duplicación entre soportes, es igual al producto de los cocientes, entre las audiencias acumuladas y las audiencias individuales, para cada soporte:

$$\frac{D_{ij}^{n_i, n_j}}{D_{ij}} = \frac{R_{n_i}}{A_i} \cdot \frac{R_{n_j}}{A_j}$$

de donde se puede despejar:

$$D_{ij}^{n_i, n_j} = \frac{D_{ij} \cdot R_{n_i} \cdot R_{n_j}}{A_i \cdot A_j}, \text{ siendo}$$

$D_{ij}^{n_i, n_j}$: duplicación entre los alcances acumulados de las n_i inserciones del soporte i y las n_j inserciones del soporte j .

D_{ij} : duplicación entre la audiencia del soporte i y del soporte j .

R_{n_i} : alcance acumulado de n_i inserciones en el soporte i .

R_{n_j} : alcance acumulado de n_j inserciones en el soporte j .

A_i : audiencia de una inserción en el soporte i .

A_j : audiencia de una inserción en el soporte j .

¹⁵⁹ MORGENSZTERN, Armand: "Frequentation de la Presse, Generalisation, Effet Publicitaire". *IREP*. Etude N° 19. Abril 1970.

El alcance neto acumulado (C_{n_i, n_j}) es simplemente la suma de los alcances acumulados por cada soporte en las n_i y n_j inserciones, menos la duplicación neta acumulada:

$$C_{n_i, n_j} = R_{n_i} + R_{n_j} - D_{ij}^{n_i, n_j}$$

Dicha expresión fue testada por Chandon¹⁶⁰ para las cinco revistas del estudio de Politz (1958). De 30 casos (las diez parejas posibles, con 2, 4 ó 6 inserciones por soporte), en 29 se produce una subestimación, que además crece con el número de inserciones: -3,46% de media para 2 inserciones, -7,12% para 4 inserciones y -10,41% para 6 inserciones. Estos resultados hacen dudar de la validez de la hipótesis de Morgensztern.

Cuando se generaliza a m soportes, con un total de N inserciones, Morgensztern propone la siguiente modificación de la fórmula de Hofmans:

$$C_{N, m} = \frac{\left(\sum_{i=1}^m R_{n_i} \right)^2}{\sum_{i=1}^m R_{n_i} + \sum_{j=1}^{m-1} \sum_{i=j+1}^m \frac{k_{ij} \cdot D_{ij} \cdot R_{n_i} \cdot R_{n_j}}{A_i \cdot A_j}}$$

Esta fórmula incorpora la hipótesis anterior y, por tanto, está suponiendo que las duplicaciones acumuladas *intra* soportes son directamente proporcionales a las duplicaciones *entre* soportes. Además requiere el conocimiento del alcance acumulado final, de cada soporte: si no se dispone de esos datos, se pueden calcular a partir del Modelo de Acumulación Beta Binomial, pero entonces se producen unos errores de estimación, que repercuten en la precisión de los resultados finales.

Este modelo fue probado satisfactoriamente por Morgensztern, para diferentes combinaciones de revistas semanales francesas, que fueron controladas por un panel

¹⁶⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 538.

durante ocho semanas. Igualmente Chandon¹⁶¹ lo aplicó a las revistas del estudio Politz (1958), calculando sus errores medios: un 1,2% para combinaciones de 4 inserciones en 3 soportes, el 0,8% para 3 inserciones en 4 soportes y el 0,6% para 2 inserciones en 3 soportes. También lo testó para las citadas 57 campañas, a partir de los datos del I.M.S., obteniendo un error medio absoluto del 2,8%, con 21 casos de sobrestimaciones y 36 de subestimaciones, y siendo 46 las campañas estimadas dentro del intervalo de confianza del 95% de la muestra; se concluye que los resultados son ligeramente peores que el modelo de Hofmans.

Hay que tener en cuenta que todas las campañas testadas se han aplicado a encuestas que recogían el alcance acumulado hasta la última inserción. Si hubiera que estimar ese dato según indicaba Morgensztern, el error final aumentaría.

9.2. Ejemplo

Vamos a estimar el alcance neto acumulado de la campaña del ejemplo anterior:

	R_i	R_j		D_{ij}	k_{ij}
<i>Muy Interesante</i>	309.918	452.872	<i>Muy Int. - TP</i>	55.208	1,1123
<i>Teleprograma</i>	236.902	301.273	<i>Muy Int. - Jueves</i>	47.445	1,1208
<i>El Jueves</i>	130.265	219.734	<i>TP - Jueves</i>	23.935	1,0697

Se había calculado el alcance acumulado, según los anuncios insertados en cada soporte: *Muy Interesante*, $R_3=0,4557$; *Teleprograma*, $R_4=0,242$; y *El Jueves*, $R_5=0,1996$. La suma de audiencias, medida en individuos, será igual a:

$$\sum_{i=1}^3 R_{n_i} = 1.455.001 \cdot [0,4557 + 0,242 + 0,1996] \approx 1.305.592$$

Se suman después las duplicaciones, con las hipótesis realizadas por Morgensztern:

¹⁶¹ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Págs. 541-3.

$$\sum_{i=1}^2 \sum_{j=i+1}^3 \frac{k_{ij} \cdot D_{ij} \cdot R_{n_i} \cdot R_{n_j}}{A_i \cdot A_j} = (1.455.001)^2 \cdot \left[\frac{1,1123 \cdot 55.208 \cdot 0,4557 \cdot 0,242}{309.918 \cdot 236.902} + \right. \\ \left. \frac{1,1208 \cdot 47.445 \cdot 0,4557 \cdot 0,1996}{309.918 \cdot 130.265} + \frac{1,0697 \cdot 23.935 \cdot 0,242 \cdot 0,1996}{236.902 \cdot 130.265} \right] \approx 533.787$$

Y finalmente podemos sustituir en la fórmula del alcance neto acumulado:

$$C_{15,3} = \frac{(1.305.592)^2}{1.305.592 + 533.787} \approx 926.709$$

10. Modelo de Casualidad Constante

10.1. Fundamentos teóricos

El Modelo de Casualidad Constante, recogido por Chandon, fue creado por Christopher Fry en 1973. Mantiene las hipótesis de heterogeneidad de los individuos y de los soportes, y la estacionariedad de las exposiciones. Por otra parte, la duplicación y la acumulación se pueden explicar mediante distribuciones beta binomiales.

Fry¹⁶² introduce el concepto de *casualidad*, que mide la forma en que un soporte va acumulando audiencia, según recibe más inserciones. Por coeficiente de casualidad se computa el cociente entre la audiencia exclusiva que aporta la segunda inserción y la audiencia que aportaría bajo la hipótesis de acumulación aleatoria; siendo R_1 y R_2 respectivamente los alcances acumulados de la primera y segunda inserción, así se expresa el coeficiente de casualidad c :

$$c = \frac{R_2 - R_1}{R_1 - (R_1)^2}$$

Si R_1 y R_2 se expresan en valores absolutos, en vez de en porcentajes sobre el público objetivo, la fórmula se define:

$$c = \frac{R_2 - R_1}{R_1 - \frac{(R_1)^2}{P}}$$

Empíricamente se observa que el ritmo de acumulación es menor que bajo la hipótesis de aleatoriedad. Lo que está haciendo precisamente el coeficiente c es medir el acercamiento a la aleatoriedad. Cuanto más se acerque c a 1, estaremos más cerca de esa hipótesis. Además, como ese ratio depende únicamente de R_1 y R_2 , se puede encontrar

¹⁶² FRY, Christopher. "The Concept of Casualness". Feedback. N° 2. Abril 1973.

una cercana relación con los parámetros A y B del Modelo Beta Binomial. Así, operando en el coeficiente de casualidad:

$$c = \frac{R_2 - R_1}{R_1 - (R_1)^2} = \frac{\frac{R_1 \cdot (R_2 - R_1)}{R_1}}{R_2 - R_1 + 2R_1 - R_2 - (R_1)^2} = \frac{\frac{R_1 \cdot (R_2 - R_1)}{2R_1 - R_2 - (R_1)^2} \cdot \left(1 + \frac{1 - R_1}{R_1}\right)}{\frac{R_1 \cdot (R_2 - R_1)}{2R_1 - R_2 - (R_1)^2} \cdot \left(1 + \frac{1 - R_1}{R_1}\right) + 1}$$

tomando $A = \frac{R_1 \cdot (R_2 - R_1)}{2R_1 - R_2 - (R_1)^2}$ y $B = \frac{A \cdot (1 - R_1)}{R_1}$, y sustituyendo:

$$c = \frac{A + B}{A + B + 1} ; \quad A = \frac{c \cdot R_1}{1 - c} ; \quad B = \frac{c \cdot (1 - R_1)}{1 - c}$$

La acumulación está en función de los hábitos de exposición y de la fidelidad al soporte. Si los lectores son muy fieles al soporte, nos alejamos de la hipótesis de aleatoriedad de la acumulación, y el porcentaje c se hace más pequeño; en el extremo contrario, la *casualidad* implica que no existe fidelidad en la audiencia, ni los individuos se comportan según unos hábitos de exposición a los medios.

Lo dicho hasta ahora de la acumulación, se podría extender análogamente a la duplicación. El coeficiente de casualidad de dos soportes será el cociente entre la audiencia agregada por el segundo soporte y la duplicación aleatoria. Como el coeficiente varía según el orden de agregación, se recomienda usar la audiencia media de los dos:

$$c_{ij} = \frac{AN_{ij} - \frac{A_i + A_j}{2}}{\frac{A_i + A_j}{2} - \left(\frac{A_i + A_j}{2}\right)^2} \quad \text{o bien} \quad c_{ij} (\%) = \frac{AN_{ij} - \frac{A_i + A_j}{2}}{\frac{A_i + A_j}{2} - \frac{\left(\frac{A_i + A_j}{2}\right)^2}{P}}$$

Como señala Chandon, el Modelo de Casualidad Constante no explica cómo estima las triplicaciones, cuadruplicaciones, ..., ni tampoco la distribución de contactos. Es un modelo que se encuentra bajo la propiedad intelectual de su autor, el cual no ha querido difundir más de lo aquí expuesto. El mismo Fry testó el modelo y obtuvo precisas estimaciones en las 45 primeras campañas del I.M.S. El error medio absoluto alcanzaba el 0,88% del alcance real. Había 32 casos de sobrestimaciones y 13 de subestimaciones. En todos los casos el error de estimación era menor que el intervalo de confianza de la muestra. Además la distribución de contactos estimada reproducía la forma de la real, detectando fielmente los picos.

10.2. Ejemplo

Supongamos una campaña de dos inserciones en Muy Interesante y dos en Teleprograma, con los datos anteriormente utilizados:

	R_1	R_2		D_u
<i>Muy Interesante</i>	309.918	452.872	<i>Muy Inter. - TP</i>	55.208
<i>Teleprograma</i>	236.902	301.273		

Así se halla el coeficiente de casualidad de la duplicación de revistas:

$$c_{ij} = \frac{(309.918 + 236.902 - 55.208) - \frac{309.918 + 236.902}{2}}{\frac{309.918 + 236.902}{2} - \frac{\left(\frac{309.918 + 236.902}{2}\right)^2}{1.455.001}} \approx 0,9827$$

Bajo la hipótesis de acumulación aleatoria, estimamos el alcance que se produciría con una inserción por soporte y con dos inserciones por soporte, y hallamos la diferencia:

a) una inserción por soporte:

$$1.455.001 - \frac{(1.455.001 - 309.918) \cdot (1.455.001 - 236.902)}{1.455.001} \approx 496.359$$

b) dos inserciones por soporte:

$$1.455.001 - \frac{(1.455.001 - 452.872) \cdot (1.455.001 - 301.273)}{1.455.001} \approx 660.373$$

c) incremento del alcance: $660.373 - 496.359 = 164.014$

Sin embargo, el coeficiente de casualidad indica que el alcance se incrementa realmente un 98'27% de lo supuesto bajo la hipótesis de aleatoriedad; es decir, será 161.777 ($\approx 164.014 \cdot 0,9827$). Esa magnitud deberá sumarse al alcance neto que se producía realmente entre las dos revistas, obteniendo el alcance neto acumulado de la campaña:

$$C = (309.928 + 236.902 - 55.028) + 161.777 = 653.579$$

B) MODELOS DE SIMULACIÓN

Conceptos previos y nueva perspectiva

La evaluación del alcance y la distribución de contactos de múltiples soportes e inserciones, puede seguir otro enfoque: medir el comportamiento individual para derivar el comportamiento del grupo, es decir, estimar los contactos que recibe cada individuo y luego agregar todos los individuos de la población objetivo en una distribución de contactos. Este enfoque se contrapone a los "modelos de fórmula", así llamados frecuentemente en la literatura y en el lenguaje corriente: su diferencia radica en que todo el análisis se hace a nivel agregado, midiendo siempre el comportamiento de toda la población. A mi juicio no es apropiado el nombre de "modelos de fórmula" pues hace pensar que los modelos de simulación no emplean fórmulas: en las siguientes páginas daré ocasión para comprobar que no es así.

Me parece imprescindible definir el sentido que voy a dar a la *simulación*. Coincide con el significado dado por la Real Academia de Ciencias Exactas, Físicas y Naturales, con el enfoque específico antes comentado: "Técnica numérica para estudiar el comportamiento de un modelo complicado, probabilístico o no, que incluye ciertos tipos de relaciones lógicas y matemáticas que son la abstracción del comportamiento y de la estructura del mundo real complejo que se quiere estudiar"¹⁶³. Pretende finalmente "resolver problemas complejos mediante su transformación en otros más simples"¹⁶⁴. Por tanto esta técnica no tiene como condición necesaria el uso de procedimientos aleatorios

¹⁶³ REAL ACADEMIA DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES. Voz correspondiente del *Vocabulario Científico y Técnico*. Espasa Calpe. Madrid 1990. Pág. 641.

¹⁶⁴ PULIDO SAN ROMÁN, Antonio. "Evolución y Revolución del Marketing". Asociación para el progreso de la dirección. Madrid. 1974. Pag. 22.

o pseudoaleatorios, como a veces se ha recogido en la literatura en lengua vernácula: Instituto Nacional de Publicidad¹⁶⁵ y Lamas¹⁶⁶.

No es correcta tampoco la clasificación realizada Bigné¹⁶⁷ en los capítulos X y XI de su tesis doctoral, posteriormente publicada en Esic-Market¹⁶⁸. Divide los modelos de evaluación en dos grupos, según se basen en la simulación o en procedimientos heurísticos. Todo criterio de clasificación debería separar un conjunto en elementos realmente distintos: sin embargo los agrupados bajo el nombre de "modelos de simulación" usan a veces procedimientos heurísticos y viceversa. La simulación y los procedimientos heurísticos no son excluyentes, aunque se hayan aplicado a distintas áreas de la planificación: la primera para medir la distribución de contactos y los segundos para seleccionar los medios y soportes. Son pues sistemas compatibles y frecuentemente se integran, de forma que no deberían usarse como alternativas de una clasificación.

Cuando la simulación se aplica a la planificación de medios, Sissors y Bumba¹⁶⁹ señalan que "se pretende reproducir, o simular, la probable exposición de los individuos en el mundo real, hasta tal punto que es posible predecir cómo se expondrán en el futuro. [...] contiene relativamente pocos algoritmos, o ninguno. El modelo de simulación mide los hábitos de lectura y de exposición a la TV de los individuos de una muestra de la población, produciendo datos en forma de probabilidades de exposición a esos medios".

¹⁶⁵ INSTITUTO NACIONAL DE PUBLICIDAD: "Investigación sobre técnicas de selección y planificación de medios publicitarios I: Metodologías para la selección óptima de soportes publicitarios". Madrid. 1977. Págs. 301-4.

¹⁶⁶ LAMAS ALONSO, Carlos. "Eso de las evaluaciones no lo veo claro". *Investigación y Marketing*. AEDEMO. Nº 42. Julio 1993. Pág. 60.

¹⁶⁷ BIGNÉ ALCANIZ, J. E. "El estudio de la audiencia y la planificación de medios y soportes publicitarios". Tesis doctoral no publicada. Universidad de Valencia. Fac. CC. EE. y Empresariales. Valencia, 1989. Págs. 638-748.

¹⁶⁸ BIGNÉ ALCANIZ, J. E. "Modelos y enfoques de evaluación de métodos publicitarios". *Esic-Market*. Nº 71. Enero-Marzo 1991. Pág. 97.

¹⁶⁹ SISSORS, Jack Z. and BUMBA, Lincoln: "Advertising Media Planning". NTC Business Books. NTC Publishing. 3ª Edition. 1990. Pág. 362.

No han faltado críticas para estos modelos de simulación, principalmente por la falta de verificaciones empíricas. Así Chandon¹⁷⁰ y Marc¹⁷¹ señalan que es necesario contrastar las estimaciones conseguidas por la simulación con los datos directamente observados de un panel. En esa línea está el experimento que he realizado, testando las estimaciones de los modelos TOM y MARKSEL sobre un panel (**I Parte, §D**).

Estas páginas no recogen todos los modelos de simulación, porque se alargaría innecesariamente esta tesis. Otras publicaciones hacen una recopilación más completa, pero menos exhaustiva: Instituto Nacional de Publicidad¹⁷² y Broadbent¹⁷³. Sin duda existen otros modelos no presentes en estas referencias. Además se han desarrollado muchas variantes con una estructura similar. Esta profusión se ha podido deber a que cada agencia ha intentado construir su modelo. Aquí se reúne una muestra de los más significativos.

¹⁷⁰ CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986. Pág. 461.

¹⁷¹ MARC, M.: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Pág. 16.

¹⁷² INSTITUTO NACIONAL DE PUBLICIDAD: "Investigación sobre técnicas de selección y planificación de medios publicitarios II: Recopilación de modelos de planificación de medios". Madrid. 1977. Págs. 301-4.

¹⁷³ BROADBENT, Simon: "Media Planning and Computers by 1970: A Review of the Use of Mathematical Models in Media Planning". *Applied Statistics*. Noviembre 1966. Vol. 15. Págs. 234-56.

1. Modelo CAM

El Modelo de Simulación CAM (*Computer Assessment of Media*) fue desarrollado por Beale, Hugles y Broadbent¹⁷⁴, para la London Press Exchange. Se desarrolló para medios impresos y TV, los principales en el Reino Unido. Los autores señalan que el CAM no trata de construir la campaña, sino su análisis: el resultado no es una campaña, sino la evaluación de una serie de campañas alternativas, por las que el planificador deberá optar.

Los datos de los medios impresos provienen de la encuesta continua *National Readership Survey* (NRS). Cada año se hacen 17.600 entrevistas, sobre 90 publicaciones. Además de los aspectos sociodemográficos, se pregunta por el equipamiento del hogar y los hábitos de consumo de algunos productos. Una persona se considera dentro de la audiencia de un número de una publicación cuando "ha visto alguna parte de ese número".

Los datos de TV provienen de la *Television Audience Measurement Ltd.* (TAM): en cada hogar, un aparato adosado al televisor registra cuándo está apagado o encendido y un diario conserva quién estaba presente en la habitación durante cada cuarto de hora (al menos 8 de los 15 minutos): cruzando la información del aparato y del diario, se obtiene la audiencia aproximada de la TV por tramos de 15 minutos. De esas exposiciones se saca la frecuencia de cada individuo (por ejemplo, la persona que ha visto 5 cuartos de hora, durante los tramos de 8.30 p.m. - 9.00 p.m. de los últimos cuatro sábados, tiene una probabilidad de exposición de 5/8). Como audiencia también se cuenta los que estén en otra habitación o hagan otra actividad dentro de la habitación: por tanto, esa audiencia ha de ser corregida por un *valor de atención*.

¹⁷⁴ BEALE, E.M.L.; HUGHES, P.A.B. and BROADBENT, S.R.: "A Computer Assessment of Media Schedules". *Operational Research Quarterly*. Vol. 17. Nº 4. Diciembre 1966. Págs. 381-411.

Las audiencias de TV y de medios impresos provienen de distintas muestras. En principio no se podría evaluar una campaña mixta, salvo que los datos se fundan en una sola base de datos, en la que cada individuo tenga simultáneamente unas probabilidades de exposición a la TV y a los medios impresos. Como ejemplifica Gensch¹⁷⁵, el problema tiene fácil solución matemática, sencillamente minimizando las distancias entre las características demográficas de los encuestados, a través de la programación lineal; en la **Figura 4.2** hemos recogido el caso propuesto por Gensch:

TELEVISION AUDIENCE MEASUREMENT					NATIONAL READERSHIP SURVEY				
Individuo	Características demográficas			Hábito de exposición	Individuo	Características demográficas			Hábito de lectura
1	1	4	5	112	6	1	3	1	214
2	2	3	2	114	7	1	4	5	863
3	1	2	1	000	8	3	6	2	110
4	2	1	4	210	9	2	1	5	002
5	3	6	2	101	10	2	3	2	415

Individuos fusionados	Hábito de exposición	Hábito de lectura
1 - 7	112	863
2 - 10	114	415
3 - 6	000	214
4 - 9	210	002
5 - 8	101	110

Figura 4.2: Fusión de datos de distintas muestras. Fuente: Gensch (op.cit.)

Ahora bien, se presentan varios problemas: NRS y TAM no tienen el mismo tamaño muestral, ni la misma definición, ni la misma estratificación. Por ello, la encuesta menor, TAM, debe adaptarse al tamaño de NRS; igualmente, como NRS no incluye los niños, la fusión con TAM tampoco debe tenerlos en cuenta. La fusión se realiza en función de la proximidad de varias características demográficas (sexo, edad, con/sin hijos, y clase social) pero también del hábito de exposición a la TV, cuestión preguntada en la encuesta del NRS.

¹⁷⁵ GENSCHE, Dennis H.: "Computer Models in Advertising Media Selection". *Journal of Marketing Research*. Noviembre 1968. Vol 5. Págs 421-3.

Los individuos deben ser ponderados en función de su importancia en el *target*. El mayor peso asignado es 100. Las ponderaciones se aplican de tal forma que dos individuos con un peso de 50 equivalen a un individuo con 100. Las ponderaciones se pueden hacer también a grupos de individuos, conocidos como "grupos característicos".

Antes se decía que la audiencia de una publicación está constituida por los que "han visto alguna parte de ese número". Ese valor debe ser corregido por un *valor de percepción*, que mide la proporción de las páginas leídas, así como la percepción subliminal, el contenido de los anuncios o el público objetivo: el resultado es la probabilidad de ver un anuncio. Pero no todos los anuncios vistos tienen el mismo valor: hay que tener en cuenta la *selectividad*, que se define como el contexto o contenido editorial, que envuelve la publicación; y el *impacto*, que es el efecto producido por el formato del anuncio. El *valor de percepción*, la *selectividad* y el *impacto* tienen como máximo valor 100 (valor que también se les da cuando se quieren obviar) y se pueden expresar en porcentaje. Los autores definen la Probabilidad de Recibir un Impresión (PRI: Probability of Receiving an Impression):

$$\text{PRI} = \text{probab. exposición a la publicación} * \text{valor de percepción} * \text{selectividad} * \text{impacto}$$

Análogamente la probabilidad de exposición a la TV ha de ser corregida por el ya comentado *valor de atención*, pero también por la *selectividad* (día de la semana y momento del día) y por el *impacto* (duración del anuncio). Finalmente se define la Probabilidad de Recibir una Impresión en TV:

$$\text{PRI} = \text{probabilidad de exposición a la TV} * \text{valor de atención} * \text{selectividad} * \text{impacto}$$

Más relevante para la toma de decisiones, es el valor de las impresiones por unidad monetaria, que incluye la ponderación de los individuos y el coste relativo. Los soportes (una revista, un periódico, una franja horaria de TV en un día de la semana, etc)

se evalúan en función de sus VIP's, y el programa los ordena en un ranking. Así se define VIP's (Valued Impressions per Pound):

$$\text{VIP's (medios impresos)} = \text{lectores ponderados} * \text{valor de percepción} * \text{selectividad} * \frac{\text{impacto}}{\text{coste en libras}}$$

$$\text{VIP's (TV)} = \text{telespectadores ponderados} * \text{valor de atención} * \text{selectividad} * \frac{\text{impacto}}{\text{coste en libras}}$$

El modelo CAM también incluye la *función de respuesta*, que evalúa la utilidad relativa para un individuo de la primera impresión, de la segunda impresión, etc. Conviene que la función de respuesta se exprese en forma acumulada. Es muy difícil de medir en la práctica. Los autores no proponen una determinada forma matemática (función lineal, función geométrica, la clásica en forma de S, definida solamente para los primeros valores,...) sino que el planificador debe decidir en cada caso, asignando una valoración a cada impresión. Este análisis se hace en relación a los objetivos de la campaña: conocimiento de marca, incitación a la compra, etc. Cuando la distribución de impresiones se pondera por los valores de la función de respuesta, llegamos a la *efectividad* de la campaña.

La suma ponderada de las particulares distribuciones de impresiones, multiplicada por la función de respuesta acumulada, da un índice de la efectividad de la campaña, que luego se relaciona con los costes. Con esta información el programa permite analizar hasta nueve campañas a la vez. Sobre una campaña se pueden añadir o suprimir soportes, y el programa reevalúa la efectividad por unidad monetaria.

Ahora nos vamos a referir a los algoritmos empleados para determinar el alcance y la distribución de contactos. Como modelo de simulación, emplea las probabilidades individuales de exposición, individuo a individuo, en vez de datos agregados. La *acumulación* sigue una sencilla distribución binomial: sea un individuo con una probabilidad de exposición p , a un determinado soporte r , en el que se hacen n , inserciones; la probabilidad de que ese individuo reciba k , de las n , impresiones vale:

$$E_{k_r}^{n_r} = \binom{n_r}{k_r} \cdot (p_r)^{k_r} \cdot (1 - p_r)^{n_r - k_r}$$

Nótese que la probabilidad p_r es la que antes se llamaba PRI (Probabilidad de Recibir una Impresión). En el caso de varios soportes o de *acumulación neta*, la probabilidad de que ese individuo reciba K impresiones entre las N inserciones de todos los soportes, vale:

$$E_K^N = \sum_{k_r=0}^K E_{K-k_{r-1}}^{N-n_{r-1}} \cdot E_{k_r}^{n_r}$$

siendo $N = \sum_{r=0}^m n_r$ la suma de las inserciones en los m soportes.

Los autores comentan que se vieron obligados a simplificar el proceso, pues los cálculos se hacían excesivamente grandes, sobre todo cuando el número de inserciones era elevado. La simplificación consistía en agrupar las PRI en intervalos.

2. Modelo SCAL

El Modelo de Simulación SCAL (*Simulation des Comportements Aleatoires de Lecture*) fue desarrollado en Francia para la empresa SEMA, y se recoge principalmente en las publicaciones de Adam-Bertier¹⁷⁶ y Bertier-Jeu¹⁷⁷. Al principio se aplicaba sólo al medio Prensa, aunque otros autores lo extienden a las Revistas (Broadbent¹⁷⁸ y Marc¹⁷⁹). Su fuente originaria era un panel de lectura, pero también se podría aplicar sobre las probabilidades de lectura, extraídas por ejemplo de una encuesta. Para Marc, el panel determina que el SCAL simule fielmente la realidad. Se trata de un modelo de evaluación, aunque aporta unos mecanismos para seleccionar el "mejor" plan.

Se establecen dos categorías de objetivos en la publicidad: objetivos de *comunicación* (contactos de los individuos ante los soportes) y objetivos de *transformación* (comportamiento de los individuos ante los productos). Los primeros se concretan en la distribución de contactos y los segundos en la curva de eficacia. La curva de eficacia mide la relación entre los cambios de comportamiento de los individuos hacia el producto y la distribución de contactos. La variación del comportamiento sigue a la modificación del conocimiento del producto y de la actitud de compra.

Este modelo se fundamenta en el *patrón de lectura*, que representa el comportamiento individual ante los diversos soportes. Dicho concepto se podría representar en tres dimensiones: el título del soporte, la fecha de aparición del soporte y la fecha de lectura del soporte. Obviando la fecha de lectura, el problema se simplifica y podría representarse sólo en dos dimensiones, como aparece en la **Tabla 4.8**; X_{it} vale 1 si

¹⁷⁶ ADAM, C. y BERTIER, P. "SCAL: un Modèle d'Evaluation des Plan de Campagne de Publicité dans le Presse". *Gestion*. Marzo 1966. Págs. 162-73.

¹⁷⁷ BERTIER, Patrice y JEU, Patrick du: "Simulation des Comportements Aléatoires de Lecture: Le Modèle SCAL". *Metra*. Vol. 6. Nº 4. 1967. Págs. 647-59.

¹⁷⁸ BROADBENT, Simon: "Media Planning and Computers by 1970: A Review of the Use of Mathematical Models in Media Planning". *Applied Statistics*. Noviembre 1966. Vol. 15. Págs. 252-3.

¹⁷⁹ MARC, M.: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Págs. 11-6.

el individuo contacta con el soporte i en periodo t , y vale 0 si no contacta; p_i es la probabilidad asignada al soporte i (esa serie de probabilidades forman el vector P).

Soportes	Periodos de tiempo						Probabilidad del soporte
	1	2	...	t	...	T	
1	X_{11}	X_{12}				X_{1T}	p_1
2	X_{21}	X_{22}				X_{2T}	p_2
...							...
i				X_{it}			p_i
...							...
S	X_{s1}	X_{s2}				X_{sT}	p_s

Tabla 4.8: Comportamiento de lectura de un individuo.

Como indica su nombre, el modelo se basa en simular aleatoriamente los comportamientos de lectura, a partir de un panel: de esa información consigue determinar la "audiencia simulada" de cualquier combinación de soportes y ejemplares.

Suponiendo que $n = \sum_{i=1}^S \sum_{t=1}^T X_{it}$ es el nº de exposiciones que recibirá un individuo, de S soportes en T periodos de tiempo, n se simula a partir del vector P , soporte por soporte, y asumiendo la hipótesis de independencia entre soportes. Como señala Marc¹⁸⁰, n puede tomar muchos valores, por la misma razón que tirar 60 dados al mismo tiempo no necesariamente produce 10 veces el seis, aunque en cada tirada la probabilidad del suceso favorable vale $\frac{1}{6}$; si repetitivamente se tiran los 60 dados, resulta que la media de apariciones del seis está muy próxima a 10, y la varianza será grande. Bertier y Jeu, en el artículo citado, indican los valores de la media y de la varianza de la variable n , que se distribuye según una normal $N(m, \sigma)$:

$$m = \Pi \cdot T \quad \text{la media de la variable, siendo } \Pi = \sum_{i=1}^S p_i$$

$$\sigma^2 = T^{1 - (\frac{1}{6})^\alpha} \cdot \left[\lambda \Pi (1 - \frac{1}{6}) + \mu \sum_{i=1}^S p_i (1 - p_i) \right]$$

¹⁸⁰ MARC, M.: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Pág. 15.

V es el n° de soportes cuya $p_i \neq 0$ ($V \leq S$). A partir de los datos, estimaron los parámetros, resultando que $\lambda=0$, $\mu=1$ y $\alpha=3$. Si se admite la hipótesis de independencia, n se distribuiría según una normal de varianza:

$$\sigma^2 = T \cdot \sum_{i=1}^S p_i (1 - p_i) \quad \text{que corresponde a los valores } \lambda=0, \mu=1 \text{ y } \alpha=\infty.$$

El análisis se extiende a F individuos; cada cual recibe n_1, n_2, \dots, n_F exposiciones de todos los soportes. Si se suman los periodos de tiempo, se puede definir la variable y_{if} que mide las veces que el individuo f se expone al soporte i . La información se podría recoger como en la **Tabla 4.9**, siendo $y_i = \sum_{t=1}^T X_{it}$ y $n_f = \sum_{i=1}^S y_{if}$.

Soportes	Individuos					
	1	2	...	f	...	F
1	y_{11}	y_{12}				y_{1F}
2	y_{21}	y_{22}				y_{2F}
...						
i				y_{if}		
...						
S	y_{S1}	y_{S2}				y_{SF}
Contactos recibidos	n_1	n_2		n_f		n_F

Tabla 4.9: Comportamiento de lectura de la muestra.

La simulación ha de someterse a varias restricciones que limitan las hipótesis de duplicación y acumulación aleatorias o independientes. Esto se acerca más a la realidad, pues las audiencias están correlacionadas positivamente, tanto *entre* como *intra* soportes. La primera restricción de la simulación es que las varianzas de las filas (soportes) y de las columnas (periodos de tiempo) deben ser menores que las que se alcanzarían respectivamente con la duplicación y la acumulación aleatorias. La segunda implica que el número de exposiciones en cada fila sea proporcional a la probabilidad de exposición del soporte: p_1/π , p_2/π , p_3/π , ..., p_S/π . El mecanismo de simulación procede

rellenando progresivamente cada línea, de acuerdo con proporciones correspondientes, y teniendo en cuenta los resultados de las líneas precedentes.

Bertier y Jeu validaron el modelo con un fichero de *patrones de lectura* durante un periodo de ocho semanas. Para 14 campañas, se comparaba la distribución de contactos simulada con la recontada directamente en el panel. Señalan que los resultados fueron muy satisfactorios, pero desgraciadamente no se recogen en el artículo. El SCAL tenía tres programas disponibles en ese momento:

- SCAL 30: proporciona la distribución de contactos de la población total y, a la vez, según el criterio seleccionado: edad, categoría socio-profesional, posesión de TV, etc.

- SCAL 31: como el anterior pero los grupos se pueden ponderar, según unos intervalos que indican su importancia para el anunciante; suministra también la distribución de contactos de la población útil (los que están dentro del público objetivo).

- SCAL 32: permite además la ponderación de los individuos, pero no uno a uno, sino por segmentos (con un máximo de 11). Presenta distribuciones de contactos para cada segmento, para el total y para la población útil.

Inicialmente clasifica los soportes según la cobertura y la rentabilidad; el usuario puede hacer planes alternativos, según los objetivos y el presupuesto. El algoritmo proporciona la distribución de contactos correspondiente. A cada intensidad de contacto se le asigna un nivel de eficacia; multiplicando los niveles por los contactos se llega a la eficacia global del plan. El modelo no va buscando el óptimo, sino que evalúa los planes propuestos en función de muchos criterios.

3. Modelo Ad-Me-Sim

El Modelo Ad-Me-Sim¹⁸¹ (*Advertising Media Simulation*) fue desarrollado en Estados Unidos por Gensch en colaboración con la agencia J. Walter Thompson. Constituye un sistema para planificar los medios y soportes que mejor difunden un mensaje publicitario determinado. Abarca un largo proceso de decisiones, con objetivos concretos, mucho más completo que la sencilla formulación del alcance o frecuencias, de anteriores modelos. Incluye factores cualitativos relevantes en la planificación. Gensch señala que la *simulación* define la forma de llegar al *target* con los costes más eficientes, con la frecuencia deseada; dicho sentido es muy distinto del que aquí se emplea, sin embargo, en ambas concepciones, éste es un modelo de simulación. El proceso tiene tres etapas:

1. Generación de los datos a través de encuestas pertinentes, en las que un mismo entrevistado debe contestar sobre sus hábitos de exposición a todos los soportes estudiados, en ese caso sólo del medio TV y revistas. Dos institutos, Brand Rating Research Incorporated (BRI) y W. R. Simmons Associates, aportaban esa información en USA, con muestras de 16.000 y 20.000 individuos respectivamente. Junto a los hábitos de audiencia, recoge las características sociodemográficas y algunas cuestiones sobre el consumo, el equipamiento, etc. Los datos del BRI, utilizados por Gensch, tenían información acerca de 200 cadenas de TV y 75 revistas.

Para predecir la futura exposición o no exposición a un soporte, Gensch propone aplicar el método de Monte Carlo sobre la probabilidad de exposición, recogida en la última encuesta del BRI. Por ejemplo, una persona que hubiera visto tal revista semanal una vez durante las últimas cuatro semanas, tendría una probabilidad del 25%; su comportamiento futuro se hace depender de números aleatorios: se generan

¹⁸¹ GENSCH, Dennis H.: "A Computer Simulation Model for Selecting Advertising Schedules". *Journal of Marketing Research*. Mayo 1969. Vol 6. Págs 203-14.

aleatoriamente números de dos dígitos (desde 00 a 99) y se van asignando a las semanas, de manera que cuando el número está entre 00 y 24, se da la exposición, o si está entre 25 y 99, no hay exposición. Sobre esas predicciones de audiencia se van a calcular todas las coberturas y frecuencias.

Las encuestas del BRI se realizaron en dos meses pero se usan para predecir (mediante el método de Monte Carlo) las audiencias del todo el año siguiente. Esto puede generar errores, sobre todo en el medio TV, donde no se pueden predecir los nuevos programas, la calidad de los programas, etc. Estas deficiencias no se deben a una mala fundamentación teórica, sino a la no disponibilidad de mejores datos en ese momento.

2. Ponderación de los individuos alcanzados por los medios: el anunciante y la agencia deben establecer la importancia de cada individuo, en función de las características de la campaña, del producto, etc. Por ejemplo, para una campaña de comida para perros, se ponderará con 1 al que tenga perro y compre diariamente el producto; en el otro extremo, se pondera con 0'001 al que no tiene perro ni compra el producto. Ni explica ni justifica por qué la escala no acaba en 0, de modo que todo entrevistado se tiene en cuenta, aunque no se incluya realmente en el *target*. La ponderación se aplica según cualquiera de las preguntas de la entrevista (véase 1). El programa Ad-Me-Sim permite también ponderar simultáneamente según distintas variables, mediante un índice geométrico ponderado.

3. Evaluación de los medios y soportes: se realizan combinaciones alternativas, debiendo tenerse en cuenta:

- Idoneidad del medio y soporte ante el mensaje publicitario (tipo de programa, contenido editorial, etc).

- Evaluación de los tipos de anuncios disponibles (por ejemplo, en TV, duración; en revistas, color o blanco y negro).

- Evaluación de las distribuciones de frecuencia ponderadas.
- Descuentos en la contratación de medios y soportes.

Los criterios de evaluación de una campaña son, para el Modelo Ad-Me-Sim:

- a) el *alcance de los medios*: la suma de individuos que reciben al menos un contacto, ponderados por su peso en el target.
- b) la *frecuencia de los medios*: la distribución de contactos, como suele entenderse, igualmente ponderados los individuos.
- c) el *alcance de los anuncios*: lo mismo que a) pero ponderado también por el tipo de anuncios.
- d) la *frecuencia de los anuncios*: igual que b) pero ponderado también por el tipo de anuncios.
- e) los *impactos publicitarios*, medido por la interacción entre los medios y la ponderación de individuos, del tipo de anuncio, de la idoneidad de los medios y de los contactos recibidos.

La planificación es un proceso continuo: no se hace en el momento inicial de la campaña y luego se abandona, sino que cada semana debe evaluarse y recomponerse. Cuando se miden esos alcances y frecuencias para una semana determinada, se denominan *estáticos* y, por otra parte, cuando se refieren a todas las semanas anteriores desde el inicio de la campaña, se llaman *acumuladas*. Por ejemplo, tenemos el alcance estático de tal semana y la distribución de contactos acumulada hasta tal semana, tanto para *medios* como para *anuncios*.

El anunciante y la agencia pueden hacer distintos planes alternativos, valorándolos según los criterios de evaluación, en relación a la rentabilidad. Las alternativas se jerarquizan en función del *alcance de los medios/anuncios* por unidad monetaria gastada, la *frecuencia de los medios/anuncios* por unidad monetaria gastada,

etc. Es frecuente encontrar que las alternativas se escalafonan de manera distinta según el criterio de evaluación. Como otros aspectos del modelo, la elección queda entonces a juicio del planificador.

Otra opción es utilizar un **programa heurístico**. Para un periodo de tiempo, por ejemplo un mes, el ordenador hace un ranking de todos los soportes, en función de los *impactos publicitarios* por unidad monetaria. En el primer soporte, se colocan las inserciones admisibles en un mes; lo mismo en el segundo; y así sucesivamente hasta que se agota el presupuesto. Aparte de las restricciones temporales, el planificador podría hacer otras que le interesasen. Como resultado se obtiene una campaña inicial. Después el programa va haciendo otras campañas, quitando los dos soportes más bajos en el ranking, y sustituyéndolos por los dos más altos, anteriormente no seleccionados. Repite este proceso hasta que en tres campañas consecutivas no se mejora el anterior coeficiente de impactos publicitarios por unidad monetaria gastada.

4. Modelo de Friedman

Friedman¹⁸² pretendía desarrollar un modelo de simulación que se adaptara a la realidad de Estados Unidos. Los datos provenían del Magazine Audience Report¹⁸³, que Simmons realizaba mediante encuesta personal a 15.033 adultos, representativos de los 124 millones de adultos que vivían en la zona continental de USA. Cada encuestado era entrevistado dos veces, para medir la repetición de audiencia, salvo 2.901 individuos que sólo respondieron una vez, estimando que su segunda exposición seguiría la tendencia general: de ahí que no todo el estudio tenga origen en observaciones empíricas. Recoge la estimación de la audiencia acumulada por cada revista, durante 3, 4, ..., hasta 12 periodos consecutivos, por sexo y lugar de lectura. También estima, a través de una distribución Beta, la distribución de contactos de 3 ó 4 inserciones en cada revista, para bastantes características demográficas.

El objetivo de Friedman es comparar las estimaciones de Simmons con los resultados obtenidos por simulación. Afirma que "si el alcance y la frecuencia generados por el modelo de simulación, para múltiples soportes e inserciones, son verosímiles y lógicos (si se aproximan a aquellos que, un entendido y hábil analista de medios, esperaría alcanzar con el uso del mejor método de cálculo existente) entonces el modelo debe considerarse capaz de estimar el alcance y la frecuencia para la planificación de medios". Chandon no compartía esta opinión, argumentando que el criterio de validez, más que en la concordancia con otros modelos reconocidos, debe estar en la validez empírica del propio modelo de simulación.

El primer paso sería la simulación de la muestra, cuyas características demográficas deben representar adecuadamente el universo. Para ello se divide la

¹⁸² FRIEDMAN, Lawrence: "Constructing a Media Simulation Model". *Journal of Advertising Research*, Vol. 10. Nº 4. Agosto 1970. Págs. 33-9.

¹⁸³ SIMMONS, W.R. & ASSOCIATES RESEARCH, INC. "1969 Magazine Audience Report". Vol. 1. New York. W. R. Simmons & Associates Research, Inc.

población simulada entre las posibles alternativas demográficas. Luego el tamaño de la submuestra se halla proporcionalmente al tamaño de su correspondiente subuniverso. No es más que una afijación proporcional, como se ejemplifica en la **Tabla 4.10**, suponiendo que la muestra de 5.000 individuos representa la población estadounidense, segmentada según el sexo y tres intervalos de edad. Friedman denomina *celda* a cada subgrupo. Las celdas serán excluyentes y habrá tantas como el producto de las alternativas de cada variable demográfica.

CELDA	Universo	Proporción	Muestra simulada
Hombres; 18-34 años	26.328.084	21,23%	1.062
Hombres; 35-49 años	17.680.288	14,26%	713
Hombres; > 49 años	16.265.289	13,12%	656
Mujeres; 18-34 años	25.359.935	20,45%	1.022
Mujeres; 35-49 años	17.862.006	14,40%	720
Mujeres; > 49 años	20.504.398	16,54%	827
TOTAL	124.000.000	100,00%	5.000

Tabla 4.10: Determinación de la muestra simulada, por afijación proporcional.

Todo lo que sigue, se va haciendo para cada revista y en cada celda. Se calcula la audiencia acumulada después de una, dos, tres y cuatro inserciones, y la distribución de contactos de cuatro inserciones. Se elige una serie de cinco probabilidades de exposición ampliamente espaciadas: por ejemplo, 0, 0'04, 0'35, 0'70 y 0'99; no es una serie única, pero no deben producir resultados negativos en ninguna fase del proceso. Por medio de las ecuaciones simultáneas¹⁸⁴ (véase **II Parte, § C.3**), se calcula a cuántos individuos se asigna la probabilidad de exposición de 0, a cuántos de 0'04, ... y a cuántos de 0'99. Esas probabilidades se asignan en estricto orden, siguiendo la numeración que los individuos tienen en la celda (esa numeración se ha determinado antes aleatoriamente). Por ejemplo, si la probabilidad 0 corresponde a 100 casos, se asigna del individuo 1 al 100; si 0'04 corresponde a 50 individuos, se asigna del individuo 101 al 150; y así sucesivamente.

¹⁸⁴ SCHREIBER, Robert J.: "Probability Assignments for the Simulation of Media Reach and Frequency". *Journal of Advertising Research*. Junio 1968. Vol. 8. Nº 2. Págs. 3-8.

Si en una revista se hacen varias inserciones, un individuo recibirá tantos contactos como resulte de simular, por el método de Montecarlo, su comportamiento ante cada inserción, a partir de su probabilidad asignada. Posteriormente se suman los resultados de los individuos de la celda. Friedman vio que se subestimaba alcance acumulado y, consecuentemente, se sobrestimaba la frecuencia media.

Propone entonces una modificación, que llama *Segunda Población*. El tamaño de la muestra simulada es igual que antes (*Primera Población*). La serie de probabilidades y su correspondencia a un determinado número de individuos no varía. Ahora cambia la forma de asignar esas probabilidades: en vez de seguir el orden de numeración, la asignación es estrictamente aleatoria. Después el proceso es igual que en la Primera Población, hasta llegar al alcance acumulado y la distribución de contactos de cualquier número de inserciones en una revista, para una celda dada. En contraposición, Friedman observa que sobrestima el alcance y subestima la distribución de contactos.

Para campañas de **múltiples revistas y múltiples inserciones**, Friedman procede con este ejemplo: una campaña consta de 3 inserciones en la *revista 1*, y 2 inserciones en la *revista 2*, y se dirige a hombres, de 18 a 34 años y de rentas altas; el proceso tiene estas etapas:

1. Extraer, de la base de datos, las audiencias de esas revistas, exclusivamente para hombres. Por ejemplo, $A_1=0,3$ y $A_2=0,2$ (expresadas en porcentaje sobre el público objetivo).

2. Extraer también el alcance neto de esas revistas, igualmente para hombres: $C_{1,2}=0,37$.

3. Medir el alcance neto (\bar{C}_{12}) si la duplicación hubiera sido aleatoria (\bar{D}_{12}). Comparar la audiencia agregada por la segunda revista en una y otra situación: en el primer caso, la revista 2 añade el 0,07, que es la mitad del 0,14 añadido con la duplicación aleatoria.

$$\overline{D}_{12} = A_1 \cdot A_2 = 0,3 \cdot 0,2 = 0,06$$

$$\overline{C}_{12} = A_1 + A_2 - \overline{D}_{12} = 0,3 + 0,2 - 0,06 = 0,44$$

$$C_{12} - A_1 = 0,37 - 0,3 = 0,07 ; \quad \overline{C}_{12} - A_1 = 0,44 - 0,3 = 0,14$$

4. Buscar la audiencia acumulada de tres inserciones en la *revista 1* y dos inserciones en la *revista 2*, pero ya en el *target* específico de la campaña (hombres, de 18 a 34 años y de rentas altas). Supongamos que respectivamente son 0,4 y 0,3. Si esos datos no estuvieran disponibles en el informe Simmons, se simulan según el modelo aquí descrito: entonces, para evitar la doble y contrapuesta solución entre Primera y Segunda Población, Friedman sugiere tomar un valor medio.

5. Análogamente con la tercera etapa, se mide el alcance neto acumulado, si la duplicación fuera aleatoria. La duplicación valdría 0,12 (0,4·0,3), el alcance neto acumulado 0,58 (0,4+0,3-0,12) y lo que añade la segunda revista 0,18 (0,58-0,4). Como realmente añade sólo la mitad, el alcance neto acumulado es 0,49 (0,4+(0,18)/2).

Para campañas de tres o más revistas, sucesivamente se agrega una a todas las anteriores, que se comportan como una revista más; todo el grupo vuelve a considerarse una sola revista, a la que se agrega una nueva... En la segunda etapa hará falta la audiencia neta de varias revistas y Friedman propone hallarla por las fórmulas de Agostini, de Metheringham, u otros modelos igualmente reconocidos.

Conviene señalar varias **críticas** a este modelo, para valorarlo adecuadamente:

a) Friedman pretende simular el comportamiento individual, pero se basa en datos agregados. Ya se dijo que la simulación es una técnica que se contrapone al tratamiento agregado de la información. No obstante, la simulación de los *target* sería un excelente complemento a los modelos de fórmula, a partir de audiencias agregadas. Aunque el objetivo era bueno, el camino no resultó correcto.

b) Propone que la acumulación se mida mediante el método empírico (**III Parte, § C.3**) que atenta contra el criterio de unicidad en la solución. Aunque es consciente de ello no sugiere ningún método alternativo.

c) Igual que apunta Chandon, me parece que el modelo de simulación debe constatarse empíricamente, más que buscarse la concordancia con otros modelos. Aparte de la verificación empírica, el modelo también adolece de fundamentación teórica en muchas fases: por ejemplo, por qué la duplicación de audiencia tiene exclusivamente dos formas de ser constante, una para hombres y otra para mujeres. Por otra parte, presenta como generalizable que la *Primera Población* subestime el alcance y sobrestime la frecuencia, y viceversa con la *Segunda Población*, pero esas afirmaciones no se basan en unos tests verificables o, si lo fuera así, el autor no lo refiere.

5. Modelo ADMOD

Desarrollado por Aaker¹⁸⁵, el ADMOD es un modelo que facilita la decisión del presupuesto oportuno, la elección entre varias formas creativas y la selección del mejor plan de medios; todo ello en función de los objetivos marcados por el anunciante. Integra, por tanto, tres áreas que hasta el momento se habían estudiado por separado, entre otras razones, porque las decisiones dependían de diferentes personas. Sin embargo, anteriores investigaciones mostraban interrelaciones entre esas áreas, y se hacía necesario un modelo integrador.

Aaker concibe el modelo publicitario desde una perspectiva más amplia que hasta ahora. No se trata tanto de evaluar el nivel de exposiciones, la curva de respuesta, la función del olvido o la consiguiente efectividad. El enfoque del ADMOD se basa, en cambio, en el comportamiento del consumidor; todo el proceso publicitario debe concebirse en función de los objetivos: la primera compra, la mejora del conocimiento de una marca o de sus atributos, etc. Por tanto, además de medir individualmente las probabilidades de exposición a los medios, es necesario conocer el comportamiento individual de compra, que también será estimado en el futuro. Este tratamiento individualizado de la información siempre es más útil que a nivel agregado, pero la utilidad es proporcional al coste.

La evaluación de un plan se hace individualmente, midiendo el impacto recibido por cada integrante de la muestra. El impacto depende de tres factores: el valor neto del cambio de conocimiento o decisión, el número y la fuente de las exposiciones, y el impacto de las exposiciones sobre la probabilidad de obtener el deseado cambio de conocimiento o decisión (la función de repetición). Luego los resultados se proyectan sobre la población real, y se calcula el beneficio esperado.

¹⁸⁵ AAKER, David A. "ADMOD: An Advertising Decision Model". *Journal of Marketing Research*. Febrero 1975. Vol. 12. Págs. 37-45.

Se requiere una muestra para cada segmento. No hace falta una muestra representativa de toda la población. El tamaño de las muestras no debe ser obligatoriamente proporcional al tamaño de los segmentos del universo. El único requisito es que cada segmento esté representado por un número de individuos razonable (por ejemplo de 100 a 200). La muestra debería ser representativa y lo ideal es seleccionar aleatoriamente a los individuos. Si la audiencia de los medios es pequeña, se hace aconsejable incrementar el tamaño muestral.

Se supone que el objetivo del anunciante es acelerar el cambio de conocimiento o comportamiento en los consumidores. Ese cambio no es inmediato, así que las decisiones se han de tomar sobre la proyección presente del comportamiento futuro de compra. Los componentes del modelo toman distinta forma según se busque ganar cuota de mercado, dar un nuevo uso al producto, mejorar la imagen de marca, etc.

Sea b_{ij} la probabilidad de que el individuo i esté expuesto al soporte j (exposición no sólo en el sentido de contactar con un miembro del *target*, sino de enlazar con el objetivo de la campaña). Esa probabilidad se supone constante en el periodo de duración de la campaña. Sea también h_{cj} la probabilidad condicional de que cualquiera de los expuestos al soporte j se exponga a la forma creativa c (en último término depende del soporte y de la creatividad, puesto que algunos soportes penetran más que otros en su audiencia, y algunos anuncios captan mejor la atención que otros). El producto de dichas probabilidades, p_{cij} , es la probabilidad de que un individuo i esté expuesto al soporte j , con la forma creativa c :

$$p_{cij} = b_{ij} \cdot h_{cj}$$

El modelo ADMOD no busca el número exacto de exposiciones recibidas por el individuo. Se llama z_i a la variable aleatoria que indica el número de exposiciones recibidas por el individuo i , y tiene asociada la distribución de probabilidades $f_{ci}(z_i)$. Como se calcula individuo a individuo, Aaker señala que aquí se puede usar una

distribución unimodal, pero cuando los individuos se agregan es inevitable usar una distribución multimodal. Se inclina por la distribución binomial, por su flexibilidad y facilidad para determinar los parámetros. La binomial se caracteriza por tener una media (μ) y una varianza (σ^2) que en este caso se expresan:

$$\mu = \sum_j p_{cij} \quad \sigma^2 = \sum_j p_{cij} \cdot (1 - p_{cij})$$

Los parámetros α y β de la binomial se pueden estimar a partir de la media y la varianza, siendo α menor o igual que el número de inserciones de la campaña:

$$\alpha = \frac{\mu^2}{\mu - \sigma^2} \quad \beta = \frac{\mu}{\alpha}$$

El impacto de las z_i exposiciones, recibidas por el individuo durante la campaña, en la probabilidad del cambio deseado, es la función de repetición, $a'_{cs}(z_i)$. El subíndice s indica que la función de repetición será distinta en cada segmento, pero idéntica para los componentes del segmento. Con c se indica que la función de repetición también está influida por la forma del anuncio. Una campaña para dar a conocer un producto requiere pocas exposiciones, pero crear una actitud exige muchas. Con el apóstrofe se quiere indicar que todas las exposiciones tienen la máxima efectividad de los soportes. Esta última hipótesis se relajará posteriormente. Debe tenerse cuidado para asegurar que las definiciones de h_{c_j} y z_i sean comparables, pues han sido definidas en contextos distintos y se estiman de forma diferentes.

Cada función de repetición tendrá una forma distinta. Cuando el individuo no recibe ninguna exposición, la probabilidad de respuesta deseada es mínima. Según aumentan las exposiciones, se mejora la probabilidad de respuesta aunque, superado cierto nivel, podría reducirse por el efecto del olvido o la saturación. Existe siempre un nivel máximo de respuesta, por debajo de la probabilidad unitaria: ese máximo depende de la forma de respuesta en cada segmento ante una creatividad dada. Aaker no define

una función de olvido determinada, pero sí asume que la campaña se olvida totalmente cuando acaba. Otra hipótesis es que la función de repetición sólo es sensible ante las exposiciones durante el periodo de la campaña.

La efectividad de los soportes, v_{cj} , se había supuesto máxima hasta aquí. Se consideraba que las exposiciones afectan al consumidor en las condiciones óptimas de prestigio, credibilidad, persuasión,... del soporte. Sin embargo ahora relajamos esa hipótesis y, en la escala de efectividad de 0 a 1, los soportes tendrán valores inferiores al máximo 1. La función de repetición ajustada, única para el individuo i , se nota como $a_{ci}(z_i)$, y es el resultado de multiplicar $a'_{ci}(z_i)$ por una media de las efectividades, ponderada por las probabilidades individuales de exposición:

$$a_{ci}(z_i) = a'_{ci}(z_i) \cdot \frac{\sum_j v_{cj} \cdot p_{cij}}{\sum_j p_{cij}}$$

El anunciante busca el mayor valor de la campaña, en cada individuo del segmento s . Ese valor es el producto del valor para la empresa del resultado deseado (w_s) por la probabilidad de que las exposiciones lo provoquen ($a_{ci}(z_i)$). Como se desconoce el número exacto de exposiciones, se deberá multiplicar también por la distribución de probabilidad ($f_{ci}(z_i)$). Estos factores luego aparecen en la función objetivo (V) que trata de maximizar el modelo ADMOD. Todo ello deberá calcularse para cada valor de z_i , para cada individuo i del segmento s y para cada uno de los segmentos, de distintos tamaños (véanse los sumatorios de la función objetivo). Se restarán los costes de la campaña, en función de la cantidad, precio y soportes a emplear (suponiendo que no hay descuentos). Así se formularía la función objetivo:

$$V = \sum_s \frac{N_s}{n_s} \sum_{i \in s} w_s \sum_{z_i=0}^{\infty} a_{ci}(z_i) \cdot f_{ci}(z_i) - \sum_j k_j \cdot x_{cj}$$

siendo

$c = \{1, 2, \dots, C\}$ índice de la alternativa creativa,

i = índice del individuo de la muestra,

$j = \{1, 2, \dots, J\}$ índice del soporte (en TV, medios impresos y/o radio),

s = índice del segmento de mercado,

N_s = tamaño del segmento s ,

n_s = tamaño de la muestra del segmento s ,

w_s = valor para la empresa, del cambio de conocimiento o comportamiento en un miembro del segmento de mercado s ,

z_i = número de exposiciones recibidas por el individuo i , dado un plan de soportes,

$a_{ci}(z_i)$ = la probabilidad de que la acción deseada para el consumidor ocurra, después de recibir las z_i exposiciones,

$f_{ci}(z_i)$ = la probabilidad de que un individuo i reciba exactamente z_i exposiciones, dado un plan de soportes,

k_j = coste de una inserción en el soporte j , y

x_{cj} = variable que indica con 1 la inserción en el soporte j con la alternativa creativa c , y con 0 la no inserción.

El modelo ADMOD emplea muestras más pequeñas que otros modelos de simulación, como el Ad-Me-Sim. Aaker lo justifica diciendo que no emplea el método de Monte Carlo y, por tanto, no requiere tantos individuos para contrarrestar la varianza de los números aleatorios; además, ADMOD puede evaluar más rápidamente los planes y seguir un procedimiento heurístico para seleccionar el mejor.

Se deberían incluir muchas restricciones: por ejemplo, el número máximo de inserciones que admite un soporte durante la campaña, en función de su frecuencia de aparición. Sin embargo, el modelo sólo introduce la restricción de que la campaña tiene

una única alternativa creativa en todas sus inserciones. Además esta hipótesis se podría relajar, pero entonces se hace difícil la medición de la función de repetición.

El procedimiento heurístico se basa en seleccionar aquella inserción que aporte mayor valor y sistemáticamente integrarla en el plan. El proceso no se detiene ante un límite presupuestario, sino cuando no existe ninguna inserción que incremente el valor del plan. Alternativamente el tope podría ser un determinado nivel en el ratio entre el valor añadido por una inserción marginal y el coste marginal. Debe permitir eliminar inserciones y/o soportes durante el proceso.

6. Modelo Calculus-Mediac

Little y Lodish¹⁸⁶ desarrollaron un modelo basado en procedimientos numéricos (*calculus*), concebido para ordenador, con un procedimiento heurístico que selecciona planes de medios buscando maximizar la respuesta de mercado, según los inputs suministrados por el usuario (presupuesto, objetivos,...).

El modelo se aplica a los medios TV, revistas y prensa. Little y Lodish denominan *opción del medio* (media option) a cada tipo de anuncio disponible en el medio: por ejemplo, 30 segundos en el Telediario, una página a color en Hola, o media página en La Vanguardia. Se hacen tres hipótesis en relación a la *opción del medio*:

- el soporte sólo se puede emplear una vez en cada momento; es decir, no puede haber dos anuncios en el mismo número de la revista o en el mismo programa de TV,
- su audiencia permanece estacionaria a lo largo del tiempo, y no tiene que ser igual a la del soporte correspondiente, y
- su formato tiene continuidad a lo largo del tiempo.

La planificación de medios, según estos autores, debe atender al menos a nueve aspectos: (1) la segmentación del mercado; (2) las ventas potenciales de cada segmento; (3) las probabilidades de exposición para cada *opción del medio* en cada segmento; (4) el coste de las inserciones; (5) el olvido de las exposiciones; (6) la estacionalidad del producto y de las audiencias de los medios; (7) la respuesta individual a la exposición; (8) la distribución de contactos; y (9) el valor de exposición en cada opción del medio.

Sean $j=\{1, 2, \dots, M\}$ opciones de medios, durante $t=\{1, 2, \dots, T\}$ periodos de tiempo, en referencia a $i=\{1, 2, \dots, S\}$ segmentos de mercado. La variable x_{jt} toma el valor 1 cuando se hace una inserción en la opción de medios j durante el periodo de

¹⁸⁶ LITTLE, John D.C. and LODISH, Leonard M.: "A Media Planning Calculus". *Operations Research*. Vol. 17. Enero-Febrero 1969. Págs 1-35.

tiempo t , y el valor 0 en caso contrario. Un individuo recibe la exposición de una inserción cuando "percibe la presencia del anuncio". La exposición o no exposición de un individuo del segmento i , a una opción del medio j , en el momento t , sigue una variable aleatoria x_{ijt} , que toma valor 1 en caso afirmativo y valor 0 en caso negativo.

Cada opción del medio puede recibir distintas ponderaciones, y el valor de una exposición será diferente de un segmento a otro. Así e_{ij} es el valor correspondiente a que un individuo del segmento i se exponga a la opción j . El incremento del nivel de exposición en el momento t para un individuo del segmento i , será la suma de todas las exposiciones en ese momento, ponderada por el valor de las exposiciones:

$$\sum_{j=1}^M e_{ij} \cdot z_{ijt}$$

Se supone que el efecto de la publicidad se reduce de manera constante, debido al olvido. Esa reducción sigue la constante de memorización α ($0 \leq \alpha < 1$), que indica el porcentaje del nivel de exposición que no se olvida de un periodo a otro. Llamando y_{it} al nivel de exposición de un individuo del segmento i en el momento t :

$$y_{it} = \alpha \cdot y_{i,t-1} + \sum_{j=1}^M e_{ij} \cdot z_{ijt}$$

Si se atiende a todas las inserciones desde el inicio de la campaña:

$$y_{it} = \alpha^t \cdot y_{i,0} + \sum_{s=1}^T \sum_{j=1}^M \alpha^{t-s} \cdot e_{ij} \cdot z_{ijs}$$

o, si se volviera indefinidamente hacia atrás, el nivel de exposición sería una suma ponderada de las exposiciones pasadas, con menores pesos cuanto más remotas fuesen:

$$y_{it} = \sum_{s=-\infty}^T \sum_{j=1}^M \alpha^{t-s} \cdot e_{ij} \cdot z_{ijs}$$

Cada individuo tiene unas ventas potenciales. Las ventas potenciales varían de un segmento a otro, y pueden ser estacionales. Las ventas potenciales de un anunciante en un periodo, es una función no lineal del nivel de exposición en ese periodo. La respuesta total del mercado es la suma de las respuesta individuales, para todos los segmentos y periodos de tiempo, y analíticamente se expresa:

$$R = \sum_{i=1}^S \sum_{t=1}^T n_i \cdot w_{it} \cdot E\{r(y_{it})\}$$

siendo:

n_i = número de individuos del segmento i ,

w_{it} = ponderación de las ventas potenciales de un individuo del segmento i en el periodo t ,

$r(y_{it})$ = función de respuesta: porcentaje de las ventas potenciales de un individuo ante un nivel de exposición y_{it} ,

$f_{it}(\cdot)$ = densidad de probabilidad de y_{it} ,

$w_{it} \cdot E\{r(y_{it})\}$ = media de las ventas potenciales realizadas por un individuo del segmento i en el periodo t .

Little y Lodish no proponen una única función de respuesta $r(y_{it})$; sí esperan que al aumentar el nivel de exposición aumente el porcentaje de las ventas potenciales, pero sin saber de qué forma, o si la tendencia se quiebra con altos niveles de exposición. Una posible ecuación, siempre creciente y asintótica en $r(y_{it})=1$, sería:

$$r(y_{it}) = r_0 + a \cdot \left(1 - e^{-b \cdot y_{it}}\right)$$

Las constantes r_0 , a y b deben ser no negativas. Por ejemplo, si se asignan los valores $r_0=0.1$, $a=0.85$ y $b=0.2$, se obtiene la curva de la **Figura 4.3**.

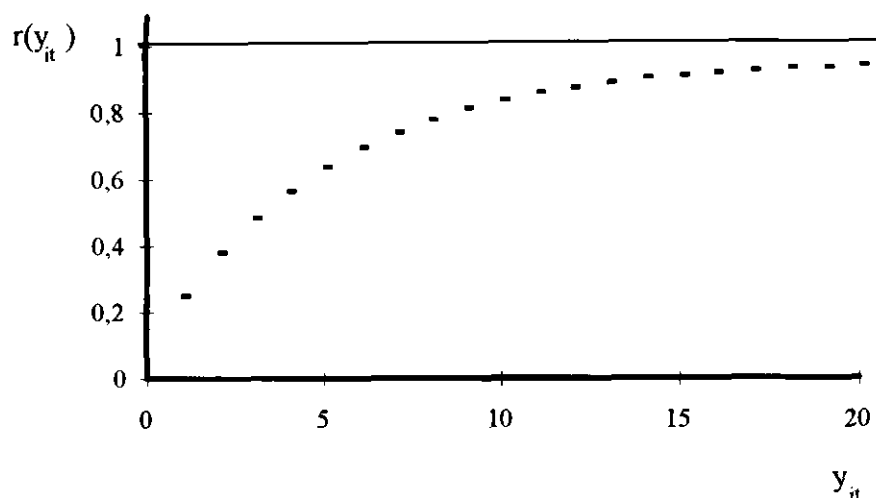


Figura 4.3: Posible función de respuesta del Modelo Calculus-Mediac.

La respuesta esperada $E\{r(y_{it})\}$ se puede expresar en los términos de los momentos de la distribución $f_{it}(y_{it})$. Eliminando los subíndices para facilitar la notación:

$$\mu = E\{y\} = \text{media de } y$$

$$\mu_n = E\{(y - \mu)^n\} = \text{momento } n \text{ de } y \text{ respecto a la media, } n > 1.$$

Al expandir $r(y)$ en una serie de Taylor respecto a la media:

$$r(y) = r(\mu) + \sum_{k=1}^{n-1} \left(\frac{1}{k!} \right) r^{(k)}(\mu) (y - \mu)^k + \left(\frac{1}{n!} \right) r^{(n)}(y_1) (y - \mu)^n$$

donde $r^{(k)}(\mu)$ es la k -ésima derivada de $r(y)$ evaluada para $y = \mu$, e y_1 es algún valor entre y y μ . Entonces:

$$E\{r\} = r(\mu) + \sum_{k=2}^{n-1} \left(\frac{1}{k!} \right) r^{(k)}(\mu) \mu_k + \left(\frac{1}{n!} \right) E\{r^{(n)}(y_1) (y - \mu)^n\}$$

En la práctica se tomarían algunos de los términos de la serie y se usaría el último para estimar el grado de aproximación. En el ejemplo de la anterior función de respuesta, se puede coger hasta el tercer momento de esta forma:

$$E\{r\} = r_0 + a(1 - e^{-b\mu}) + ae^{-b\mu} \left\{ -(\frac{1}{2})b^2\mu_2 + (\frac{1}{6})b^3\mu_3 \right\} + \varepsilon_4$$

donde

$$\varepsilon_4 = -(\frac{1}{24})aE\{e^{-by_i} b^4 (y - \mu)^4\} \text{ y}$$

$$|\varepsilon_4| \leq (\frac{1}{24})ab^4\mu_4 \quad \text{ya que} \quad e^{-by_i} = 1 \quad \text{y} \quad (y - \mu)^4 \geq 0.$$

A continuación Little y Lodish expresan los momentos de la distribución de exposiciones en términos de las x_{jt} decisiones de medios. Muestran que la media (primer momento) y la varianza (segundo momento) del nivel de exposición están en función únicamente de las probabilidades individuales de exposición y de las duplicaciones. Para simplificar, obvian momentáneamente el segmento i y el periodo de tiempo t . Sean:

y = nivel de exposición de un individuo

$$z_j = \begin{cases} 1, & \text{si el individuo está expuesto a la opción } j \\ 0, & \text{si no está expuesto} \end{cases}$$

$$y = \sum_{j=1}^M e_j \cdot z_j$$

$p_j = P(z_j=1)$ = probabilidad de que el individuo esté expuesto a la opción j

$p_{jk} = P(z_j=1; z_k=1)$ = ídem expuesto simultáneamente a las opciones j y k .

El momento de primer orden de y , o media, es simplemente:

$$\mu = E\{y\} = \sum_{j=1}^M e_j \cdot p_j$$

Luego el segundo momento μ_2 , o varianza $V(y)$, se puede expresar:

$$E\{y^2\} = E\left\{\left(\sum_{j=1}^M e_j z_j\right)^2\right\} = \sum_{j=1}^M \sum_{k=1}^M e_j e_k E\{z_j z_k\} = \sum_{k=1}^M e_j^2 p_j + 2 \sum_{j=1}^{M-1} \sum_{k=j+1}^M e_j e_k p_{jk}$$

$$\mu_2 = V(y) = E\{(y - \mu)^2\} = \sum_{k=1}^M e_j^2 p_j (1 - p_j) + 2 \sum_{j=1}^{M-1} \sum_{k=j+1}^M e_j e_k (p_{jk} - p_j p_k)$$

Los momentos de superiores ordenes exigen cálculos más complejos, pero los autores señalan que se pueden estimar a partir de los dos primeros. La probabilidad de que un individuo del segmento i esté expuesto a una inserción en el momento t , con una opción de medios j , vale:

$$p_{jit} = g_{ji} \cdot h_j \cdot s_{jt} \cdot x_{jt}$$

siendo: g_{ji} = porcentaje de individuos del segmento i que son audiencia del soporte de opción j (audiencia media del año), y

h_j = probabilidad de que una persona que es audiencia del soporte de j , esté expuesto a la opción del anuncio.

s_{jt} = índice de la estacionalidad del soporte de opción j en el periodo t (el valor medio sobre un año es 1).

x_{jt} = variable que con valor 1 indica la inserción en el soporte j durante el periodo t , y con valor 0 la no inserción.

Se está asumiendo implícitamente que la estacionalidad de la audiencia es igual en todos los segmentos, y que h_j no cambia estacionalmente. Por otra parte se asume la estacionariedad con carácter general. El modelo de Little y Lodish hace la hipótesis de que la duplicación es aleatoria, como fue suficientemente comentado (**II Parte, § B.6**); así se expresa para los soportes j y k , en los correspondiente periodos t y s :

$$p_{jk|its} = g_{jk|i} \cdot h_j \cdot h_k \cdot s_{jt} \cdot s_{ks} \cdot x_{jt} \cdot x_{ks}$$

Se debe tener en cuenta la restricción presupuestaria; siendo c_{jt} el coste de una inserción en la opción j en el periodo t , y B el presupuesto, se expresa: $\sum_{j=1}^M \sum_{t=1}^T c_{jt} \cdot x_{jt} \leq B$.

Los autores, conscientes de la desviación con la práctica presupuestaria, luego incluyen una forma de evaluación de los descuentos.

Las ventas actuales no sólo dependen de la campaña actual, sino también de la pasada. Por otra parte, la campaña actual seguirá teniendo efectos en las ventas aun cuando haya terminado. Todo ello se tiene en cuenta en el modelo, a través de E (número de periodos extra añadidos después del final de la campaña) y $K+1$ (número de periodos extra añadidos antes del inicio de la campaña).

Con todas los planteamientos anteriores, Little y Lodish plantean un programa matemático. Los efectos en las ventas futuras se miden suponiendo que la campaña continúa más allá de su fin. Se trata de maximizar la función R sujeta a la restricción presupuestaria:

$$R = \sum_{i=1}^S \sum_{t=1}^{T+E} n_i w_{it} \left\{ r(\mu_{it}) + \sum_{m=2}^n \left(\frac{1}{m!} \right) r^{(m)}(\mu_{it}) \mu_{mit} \right\} \sum_{i=1}^M \sum_{t=1}^T c_{jt} x_{jt} \leq B$$

$$\mu_{it} = \sum_{s=-K}^t \sum_{j=1}^M \alpha^{t-s} e_{ij} h_j g_{j|t} s_{js} x_{js}$$

$$\begin{aligned} \mu_{2it} = & \sum_{s=-K}^t \sum_{j=1}^M (\alpha^{t-s} e_{ij})^2 h_j g_{j|t} s_{js} x_{js} + 2 \sum_{s=-K}^{t-1} \sum_{r=s+1}^t \sum_{j=1}^M \sum_{k=1}^M \alpha^{2t-r-s} e_{ij} e_{ik} h_j h_k g_{j|t} s_{js} s_{kr} x_{js} x_{kr} + \\ & + 2 \sum_{s=-K}^t \sum_{j=1}^{M-1} \sum_{k=j+1}^M \alpha^{2t-2s} e_{ij} e_{ik} h_j h_k g_{j|t} s_{js} s_{ks} x_{jt} x_{ks} - \mu_{it}^2 \end{aligned}$$

expresando el momento m en función del primero y segundo $\mu_{mit} = \mu_m(\mu_{it}, \mu_{2it})$

$m=\{3, 4, \dots, n\}$ $i=\{1, 2, \dots, S\}$ $t=\{1, 2, \dots, T+E\}$ $j=\{1, 2, \dots, M\}$

$x_{jt} \in \{0, 1\}$ para todo (j, t) ; $x_{jt}=1$ para $(j, t) \in I_1$; $x_{jt}=0$ para $(j, t) \in I_2$

I_1 = conjunto de inserciones seleccionadas ; I_2 = conjunto de inserciones rechazadas

Los autores señalan que este programa matemático ha de desarrollarse mediante una aproximación heurística, más que con un enfoque de optimización, que ellos mismos habían probado¹⁸⁷. La solución obtenida no será entonces necesariamente óptima. El proceso heurístico se refleja en la **Figura 4.4**: se basa en escoger sucesivamente la inserción que produzca mayor respuesta, en relación con su coste; cuando se cubre el presupuesto, se asegura de que la última inserción tiene menos valor que todas las inserciones anteriormente seleccionadas.

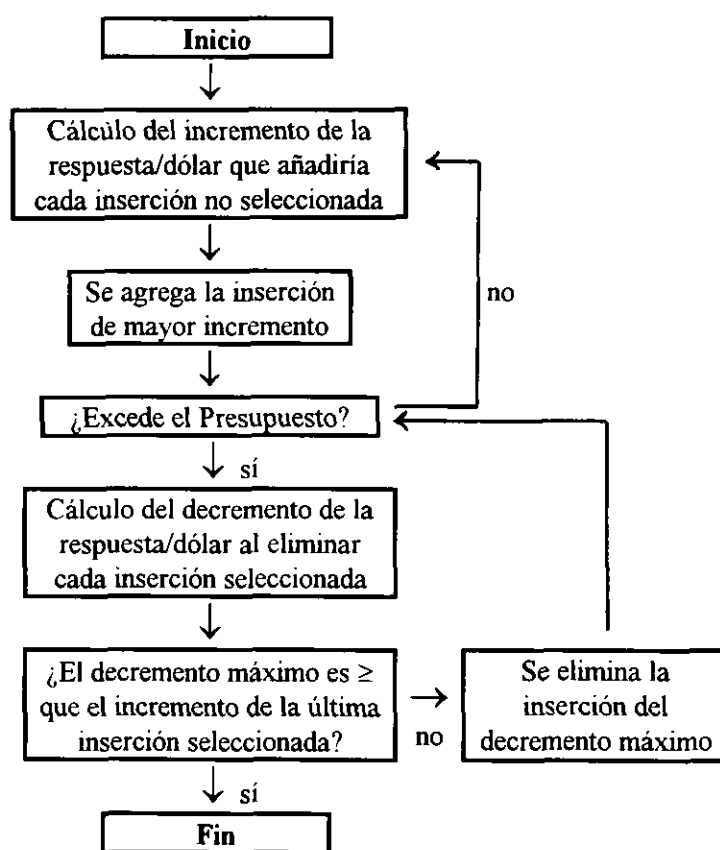


Figura 4.4: Proceso heurístico del Modelo Calculus-Mediac.

Existe un problema: el coste de las inserciones (input del proceso) depende del volumen contratado (output). Para solucionarlo, los autores proponen empezar con los menores costes (o mayores descuentos); si el proceso genera un plan con unos

¹⁸⁷ LITTLE, John D.C. and LODISH, Leonard M. "A Media Selection Model and Its Optimization by Dynamic Programming". *Industrial Management Rev.* Vol. 8, 1966. Págs. 15-24.

volúmenes que no justifican esos descuentos, se vuelve al inicio, pero dando el coste real a aquella opción con mayor discrepancia media; y sucesivamente se recomienza el proceso, hasta que el coste real de cada opción seleccionada, no supere el que tenía asignado al iniciar el proceso.

Los autores construyeron el programa MEDIAC, para usar el modelo a través del ordenador. El usuario podía acceder on-line desde su computadora y llevar una comunicación dialogada durante el proceso. Todo ello suponía un sustancial progreso. Sin embargo, el mayor plan obtenido se limitaba a 24 opciones de medios, 10 periodos de tiempo y 15 segmentos de mercado. Little y Lodish señalan que la contribución más importante es hacer más comprensible la estructura de la planificación en su conjunto.

Posteriormente Lodish¹⁸⁸ señaló que las hipótesis eran consistentes con anteriores investigaciones empíricas, referenciadas en el citado artículo: concluía que la estructura del MEDIAC se adaptaba a las observaciones sobre el recuerdo y el olvido (Zielske), el valor de exposición (Wickelgren y Norman), la constante de memorización (Simmons and Associates), la modificación de la actitud (Cromwell y Kunkel) y la función de respuesta (Rohloff). También publicó una modificación del modelo¹⁸⁹, que añade el efecto de la planificación de medios de la competencia. Admite que un individuo se puede exponer más de vez a una inserción en medios impresos (un ejemplar puede ser leído en varias ocasiones) a diferencia de los medios TV y radio. Los efectos se miden en la respuesta total del mercado, con las ventas de todos los productos. Intervienen también la eficacia relativa del marketing de cada empresa, medida por el nivel de precios y de distribución. Este ambicioso modelo intenta explicar todo el comportamiento del mercado, integrando gran cantidad de factores. Ray y Sawyer¹⁹⁰ realizaron experimentos

¹⁸⁸ LODISH, Leonard M. "Empirical Studies on Individual Response to Exposure Patterns". *Journal of Marketing Research*. Vol. 8. Mayo de 1971. Págs. 212-8.

¹⁸⁹ LODISH, Leonard M. "Considering Competition in Media Planning". *Management Science*. Vol. 17. Nº 6. Febrero 1971. Págs. B-293-B-306.

¹⁹⁰ RAY, M.L. and SAWYER, A.G. "Behavioral Measurement for Marketing Models: Estimating the Effects of Advertising Repetition for Media Planning". *Management Science*. Vol. 18. Nº 4. Diciembre 1971. Págs. P-73-P89.

de laboratorio para determinar algunos de los parámetros del modelo, aunque Boulding¹⁹¹ critica que no se pueden contrastar en la realidad.

¹⁹¹ BOULDING, W.F. "A Media Model Based on Linking Brand Choice Behaviour to Advertising Insertions". Tesis Doctoral. Universidad de Pensilvania. University Microfilms International. Michigan. Pág. 40.

CONCLUSIONES

Con *el alcance y la distribución de contactos publicitarios* hemos estudiado unos instrumentos para evaluar y comparar campañas, y cubrir un primer paso en la medición de su eficacia. Pero desde múltiples vías, evitando el reduccionismo de la "solución única". Como se anticipaba en la **Introducción**, hemos distinguido tres estadios, y a cada uno se le ofrece el instrumento adecuado para tomar decisiones más racionalmente: para pequeñas inversiones, el planificador podría servirse del modelo de fórmula más preciso, con un coste reducido o nulo; cuando el tamaño de las campañas justifica el coste de los modelos de simulación, podría optar por el algoritmo con menores errores de estimación; y si los estudios continuos se extendieran a los medios impresos, se podrían evitar los errores de estimación, además de conseguir abundante información adicional.

Los datos de la inversión publicitaria anual (**I Parte, §A**) muestran inequívocamente que muchos anunciantes no pueden permitirse evaluar sus campañas con *modelos de simulación*: en Diarios, casi 30.000 millones se reparten entre 16.000 anunciantes, y otros 40.000 millones corresponden a publicidad clasificada y local no distribuida; en Revistas, casi 20.000 millones se distribuyen entre 5.500 anunciantes; es patente la gran diferencia con los 250 primeros anunciantes de cada medio, aunque constituyan una realidad muy heterogénea; dadas sus estrecheces presupuestarias, se entiende que se deben conformar con los *modelos de fórmula*. Visto desde las agencias y centrales, las asociadas en AIMC son una pequeñísima parte del listado de la Guía de Medios de Comunicación; como se dijo, las agencias no asociadas no disponen del fichero individualizado y, por tanto, no pueden usar los modelos de simulación; y algunas de ellas, según lo declarado, facturan un importante volumen de inversión.

En la **I Parte** también se ha expuesto el funcionamiento del EGM, la fuente de información más completa y difundida en España. El modelo TOM, cuyos programas están generalizados también aquí, se ha descrito completamente, haciendo énfasis en los

algoritmos internos. Con esas fuentes de datos y esos programas, se han contrastado empíricamente los modelos de fórmula de la **II Parte**: por primera vez, al menos en la literatura, se aplican tales modelos sobre las audiencias probabilizadas del EGM, ofreciendo así instrumentos sustitutivos, utilizables a partir de los datos disponibles en España.

Las estimaciones de varios modelos se han contrastado con las directamente observadas en un panel generado aleatoriamente (**I Parte, §D**). Los *alcances* fueron ligeramente subestimados por todos los modelos, pero el TOM era más preciso que el MARKSEL y éste, a su vez, más que Metheringham. La imprecisión de todos crecía al aumentar la duración de la campaña; sin embargo al aumentar los soportes, los errores disminuyen en TOM y MARKSEL, pero crecen suavemente en Metheringham; como resultado final, TOM y MARKSEL mejoran con muchas inserciones, mientras que Metheringham mejora ante cualquier inserción adicional.

Particularmente significativos son los errores detectados en las *distribuciones de contactos*; por término medio el TOM yerra aproximadamente en el 20% de las frecuencias, Metheringham en el 10% y MARKSEL en el 8%; el primero mejora y los demás empeoran ante aumentos de inserciones o de periodos; el TOM parece indiferente ante la adición de soportes, pero Metheringham y MARKSEL reaccionan negativamente. Visto contacto a contacto, el TOM siempre subestimaba el primer y el último contacto, y tendía a sobrestimar los demás; se observaban otras tendencias, pero más suaves y menos generales: Metheringham tendía a sobrestimar la mitad central de la distribución y a subestimar el resto; por su parte, MARKSEL tendía a subestimar ligeramente todos los contactos excepto el último, que solía sobrestimarse.

Podrían hacerse dos lecturas de los resultados del TOM: admitir esos márgenes de confianza o rechazar los errores de tales magnitudes. Nuestra opinión es que hay que intentar corregirlos, porque esas estimaciones fundamentan el resto de la planificación: si

no el posterior cálculo de las curvas de respuesta, la eficacia, la optimización, etc, va multiplicando los errores y, finalmente, los resultados pueden diferir bastante de la realidad. Posteriores investigaciones deberían analizar estas desviaciones sobre paneles reales, particularizar en una mayor variedad de campañas, hacer hipótesis menos restrictivas en la generación aleatoria del panel, etc. Las partes integrantes del *segundo estadio* deberían contrastar la precisión del TOM y de otros modelos de simulación, y evaluar la viabilidad económica de un panel de Diarios y Revistas. Probablemente se produciría una mejora de todo el conjunto. Y, como apuntan otros modelos de simulación (**IV Parte, §B**) se han de fusionar las diversas fuentes continuas, de audiencia de medios y de consumo de productos, actualmente disponibles en España.

Sin repetir los resultados (**II Parte**), el modelo de fórmula que mejor estima el alcance neto es el de Agostini, con el valor de $k=1'048$ que hemos obtenido para España (valor que produce la mitad de errores que el anterior $k=1'125$). En estas páginas queda estadísticamente demostrado que k no es constante pero, asumiendo que lo fuera, la fórmula de Agostini produce errores tan pequeños que se podrían despreciar. Otros modelos con estimaciones aceptables y similares son, por orden: Kaatz (con la modificación aquí realizada), Hofmans y Metherringham. Han de ser rechazados los modelos de Starch, Kwerel, Sainsbury modificado, Kuhn, Binomial y Sainsbury, porque producen errores comparativamente muy superiores. Algunos tenían la gran ventaja de no requerir las duplicaciones de audiencia (Binomial, Sainsbury y Lozano) pero sus estimaciones eran muy imprecisas. La tendencia general era que cuantos más soportes incluyera la campaña, los modelos producían mayores errores: en consonancia con mayores inversiones, se recomienda acudir a los modelos de simulación. Se detectaba también que las estimaciones no variaban ante la forma de selección de los soportes pero sí ligeramente ante la composición del público objetivo, aunque sin tendencias definidas.

Entre los modelos de audiencia neta aceptados, el de Metherringham estima la distribución de contactos sin excesiva complejidad de cálculo. Como ventajas, este

modelo sólo requiere las duplicaciones y es extensible a la audiencia neta acumulada. El modelo Binomial no necesitaba duplicaciones pero producía errores muy superiores.

Hemos expuesto varios modelos que estiman las duplicaciones entre soportes (**II Parte, § A.6 y B.5**). Aunque algunos explican buena parte de este fenómeno, ninguno consigue la precisión que luego requiere el cálculo del alcance y la distribución de contactos. Entonces se hace necesario que las duplicaciones se publiquen periódicamente, como se referencia de otros países. Las duplicaciones de los Diarios y las Revistas asociadas al EGM no supondrían un volumen de información muy grande y se podrían recoger para varios *targets* (como actualmente se editan las audiencias para "individuos" y "amas de casa").

Los modelos de audiencia acumulada se han descrito en la **III Parte**, pero no se han verificado empíricamente, ya que el EGM es una fuente de información periódica y, por tanto, no tiene datos continuos donde observar el fenómeno de la acumulación. De las contestaciones de los entrevistados se infiere la probabilidad de audiencia individual y la audiencia probabilizada del conjunto (**III Parte, § C.5**).

Los modelos de fórmula de audiencia neta acumulada (**IV Parte, §A**), también necesitan datos, al menos de la acumulación de la segunda inserción; por eso tampoco se han verificado empíricamente sobre datos del EGM; como se comentó, el modelo de Metheringham se contrastó en el experimento de la **I Parte**, donde las campañas eran multi-periodo y multi-soporte; sobre un panel sí se pueden contar las duplicaciones *intra* y *entre*-soportes.

Los modelos de simulación, aplicados a múltiples periodos y soportes, cierran esta tesis (**IV Parte, §B**). Se explican detalladamente los más importantes recogidos en la literatura. Generalmente son programas que contienen numerosas utilidades para la planificación, y calculan el alcance y la distribución de contactos a partir de algoritmos sencillos; buena muestra son los comercializados en España: TOM y MARKSEL.

BIBLIOGRAFÍA

- AAKER, David A. "ADMOD: An Advertising Decision Model". *Journal of Marketing Research*. Febrero 1975. Vol. 12. Págs. 37-45.
- ADAM, C. y BERTIER, P. "SCAL: un Modèle d'Evaluation des Plan de Campagne de Publicité dans le Presse". *Gestion*. Marzo 1966. Págs. 162-73.
- AGOSTINI, Jean-Michel. "Analysis of Magazine Accumulative Audience". *Journal of Advertising Research*. Vol. 2. Nº 4. Diciembre 1962. Págs. 24-7.
- AGOSTINI, Jean-Michel. "How Estimate Unduplicated Audiences". *Journal of Advertising Research*. Marzo 1961. Vol 1. Nº 3. Págs. 11-14.
- ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: A Study of Duplication". Time, Inc. New York. 1953.
- ALFRED POLITZ RESEARCH, INC. "A Study of Four Media: Their Accumulative and Repeat Audiences". Time, Inc. New York. 1953.
- ALFRED POLITZ RESEARCH, INC. "Life Study of Consumer Expenditures". 7 Volumes. Time, Inc. New York. 1958.
- ALFRED POLITZ RESEARCH, Inc. "Life Study of Consumer Expenditures". 1959. New York.
- ALFRED POLITZ MEDIA STUDIES. "Main Report 1964 Magazine Study: Audience of 11 Magazines". Febrero 1965. New York.
- ALÓS, Joan S. "Industria de los Estudios de Mercado en España 1992" (Nº 44. Marzo de 1994. Pág. 104-6); "Industria de los Estudios de Mercado en España 1993" (Nº 47. Marzo de 1995. Págs. 74-6); "Industria de los Estudios de Mercado en España 1994" (Nº 48. Julio de 1995. Págs. 74-7). *Investigación y Marketing*. AEDEMO.
- ANUNCIOS REVISTA. Nº 44. Marzo de 1995.
- ASOCIACIÓN PARA LA INVESTIGACIÓN DE MEDIOS DE COMUNICACIÓN (AIMC). ESTUDIO GENERAL DE MEDIOS (EGM). Resultados Acumulados 1ª + 2ª + 3ª Olas 1991: Individuos y Hogares. Febrero - Noviembre de 1991.
- AUDITS & SURVEYS COMPANY, INC. "Look Audience Study, 1961". New York: Cowles Magazines and Broadcasting, Inc., 1961.

- BARREIRO FERNÁNDEZ, Jose Manuel. "Las duplicaciones de audiencias entre soportes publicitarios: desarrollo de un modelo informático de la fórmula de Hofmans para la determinación de coberturas y distribuciones de exposiciones de una combinación de soportes". Ponencia presentada en los III Encuentros de Profesores Universitarios de Marketing. Salamanca. Septiembre 1991.
- BEALE, E.M.L.; HUGHES, P.A.B. and BROADBENT, S.R.: "A Computer Assessment of Media Schedules". *Operational Research Quarterly*. Vol. 17. Nº 4. Diciembre 1966. Págs. 381-411.
- BERTIER, Patrice y JEU, Patrick du: "Simulation des Comportements Aléatoires de Lecture: Le Modèle SCAL". *Metra*. Vol. 6. Nº 4. 1967. Págs. 647-59.
- BIGNÉ ALCANIZ, J. E. "El estudio de la audiencia y la planificación de medios y soportes publicitarios". Tesis doctoral no publicada. Universidad de Valencia. Fac. CC. EE. y Empresariales. Valencia, 1989.
- BIGNÉ ALCANIZ, J. E. "Modelos y enfoques de evaluación de métodos publicitarios". *Esic-Market*. Nº 71. Enero-Marzo 1991.
- BOULDING, W.F. "A Media Model Based on Linking Brand Choice Behaviour to Advertising Insertions". Tesis Doctoral. Universidad de Pensilvania. University Microfilms International. Michigan.
- BOWER, John. "Net Audiences of U.S. and Canadian Magazines: Seven Test of Agostini's Formula". *Journal of Advertising Research*. Marzo, 1963. Vol 3. Nº 1. Págs. 13-20.
- BROADBENT, Simon: "Media Planning and Computers by 1970: A Review of the Use of Mathematical Models in Media Planning". *Applied Statistics*. Noviembre 1966. Vol. 15. Págs. 234-56.
- BROADBENT, Simon. "Regularity of Reading". *Journal of Marketing Research*. Agosto 1964. Págs. 50-8.
- CAFFYN, J. M. y SAGOVSKY, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods". *Journal of Advertising Research*. Marzo 1963. Vol. 3. Nº 1. Págs. 21-25.
- CANGUILHEM, Jean François. "Cómo definir prácticamente una población-objetivo publicitaria". *Información de publicidad y marketing*. Nº 119. 1973. Págs. 37-44.
- CANGUILHEM, Jean François. "Futuro de la informática aplicada a la planificación". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985.

- CANNON, Hugh M. "Reach and Frequency Estimates for Specialized Target Markets". *Journal of Advertising Research*. Vol 23. Nº 3. Junio/Julio 1983. Págs: 45-50.
- CHANDON, Jean-Louis. "A Comparative Study of Media Exposure Models". Garland Publishing, Inc. New York & London. 1986.
- CHANDON, Jean-Louis. "Le point sur les modèles de duplication d'audience". *Revue Française du Marketing*. Nº 53. 1974. Págs. 51-76.
- CHARNES, A.; COOPER, N.W.; LEARNER, D.B.L.; and SNOW, E.F.: "Media Mixing for Distribution of Frequencies with β -like and Lognormal Distributions and Frequencies". Management Science Report Nº 189. Carnegie-Mellon University. Pittsburg. Junio 1968.
- CLAYCAMP, H.J. and McCLELLAND, C.W. "Estimating Reach and the Magic of K". *Journal of Advertising Research*. Junio 1968. Vol. 8. Nº 2. Págs. 44-51.
- CLEMENTE, Jorge. "El Estudio General de Medios (I)". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios-Televisión. Gandía. Febrero 1985. Págs. 121-48.
- DÍEZ DE CASTRO, Enrique Carlos y MARTÍN ARMARIO, Enrique. "Planificación publicitaria". Ed. Pirámide. Madrid, 1993.
- DURAND, Jacques. "L'analyse de la duplication entre supports de publicité et ses enseignements pour la théorie statistique". *Gestion*. Diciembre 1965.
- DURAND, JACQUES. "L'Analyse Statistique de la Duplication". Institut de Recherches et d'Etudes Publicitaires (IREP). Nº 11. París. Noviembre 1964.
- ENGELMAN, Fred L. "An Empirical Formula for Audience Accumulation". *Journal of Advertising Research*. Vol. 5. Nº 2. Junio 1965. Págs 21-8.
- ERHENBERG, A.S.C. and GOOGHART, Gerald J.: "Practical Applications of the Duplication of Viewing Law". *Journal of Marketing Research Society*, 11. Enero de 1969. Págs. 6-24.
- FRIEDMAN, Lawrence: "Constructing a Media Simulation Model". *Journal of Advertising Research*. Vol. 10. Nº 4. Agosto 1970. Págs. 33-9.
- FRY, Christopher. "The Concept of Casualness". *Feedback*. Nº 2. Abril 1973.
- GENSCH, Dennis H.: "A Computer Simulation Model for Selecting Advertising Schedules". *Journal of Marketing Research*. Mayo 1969. Vol 6. Págs 203-14.

- GENSCH, Dennis H.: "Computer Models in Advertising Media Selection". *Journal of Marketing Research*. Noviembre 1968. Vol 5. Págs 421-3.
- GONZÁLEZ LOBO, M^a Ángeles. "Curso de Publicidad". Ed. Eresma & Celeste Ediciones. Madrid. 1994.
- GOODHARDT, Gerald J. and EHRENBURG, A.S.C.: "Duplication of Television Viewing Between and Within Channels". *Journal of Marketing Research*. Mayo de 1969. Págs. 169-78.
- GREENE, Jerome D. and STOCK, J. Stevens. "A Rate-Frequency Model of Behavior". *Journal of Advertising Research*. Vol. 11. N° 4. Agosto 1971. Págs. 9-19.
- GREENE, J. D. and STOCK, J. S. "Advertising Reach and Frequency in Magazines". Marketmath Inc. and Reader's Digest Association. New York. 1967.
- GREENE, Jerome D.: "Personal Media Probabilities". *Journal of Advertising Research*. Vol. 10. Octubre 1970. Págs. 12-8.
- GUÍA DE MEDIOS DE COMUNICACIÓN. N° 119. Septiembre-Noviembre de 1994. Ed: Remarca, S.A. Madrid.
- HEADEN, Robert S.; KLOMPMAKER, Jay E.; y RUST, Roland T. "The Duplication of Viewing Law and Television Media Schedule Evaluation". *Journal of Marketing Research*. Vol. 16. Agosto 1979. Págs. 333-40.
- HOFMANS, Pierre. "Measuring the Cumulative Net Coverage of Any Combination of Media". *Journal of Marketing Research*. Agosto de 1966. Vol. 3. Págs. 269-78.
- HOFMANS, Pierre. "Une Méthode Pratique pour calculer la Couverture Nette Cumulée d'une Série de Supports". Paris. *IREP*, Study N° 12. Junio de 1965.
- HYETT, G. P.: "The Measurement of Readership". London School of Economics Paper. London. 1958.
- HYETT, G. P.: Paper read to the Statistics Seminar. London School of Economics. Febrero de 1958.
- INSTITUTO NACIONAL DE PUBLICIDAD: "Investigación sobre Técnicas de Selección y Planificación de Medios Publicitarios I: Metodologías para la Selección Óptima de Soportes Publicitarios". Madrid. 1977.
- INSTITUTO NACIONAL DE PUBLICIDAD - EMOPÚBLICA: "Investigación sobre Técnicas de Selección y Planificación de Medios Publicitarios, II: Recopilación de Modelos de Planificación de Medios". Madrid. Mayo de 1978.

- INVERSEN, G.R. y NORPORTH, H. "Analysis of Variance". Quantitative Applications in the Social Sciences Nº 1. A Sage University Paper Series.
- J. WALTER THOMPSON. "La inversión publicitaria en España: 1990". Madrid. 1991.
- KAATZ, Ronald B. "Improving Agostini's Formula for Net Audience". *Journal of Advertising Research*. Septiembre 1963. Vol 3. Nº 3. Págs. 43-4.
- KILLION, Kevin C. "Media Models: Questions Needing Answers for Researchers Needing Questions". *Journal of Advertising Research*. Research Currents. Vol. 30. Nº 5. Octubre/Noviembre de 1990. Págs. RC-7 y 8.
- KLOTZ, J. "Statistical Inference in Bernoulli Trials with Dependence". *The Annals of Statistics*. 1973. Vol. 1. Págs. 373-9.
- KRESHEL, Peggy J.; LANCASTER, Kent M.; and TOOMEY, Margaret A. "How Leading Advertising Agencies Perceive Effective Reach and Frequency". *Journal of Advertising*. Vol. 14. Nº 3. 1985. Págs. 32-8.
- KUHN, Walter. "Net Audiences of German Magazines: A New Formula". *Journal of Advertising Research*. Marzo 1963. Vol. 3. Nº 1. Págs. 30-33.
- KWEREL, SEYMOUR M. "Estimating the Unduplicated Audience of a Combination of Media Vehicles - Integrated Theory and Estimation Method". Ph. D. Dissertation. Columbia University. 1964.
- KWEREL, SEYMOUR M. "Estimating Unduplicated Audience and Exposure Distribution". *Journal of Advertising Research*. Vol. 9. Nº 2. Junio de 1969. Págs. 46-53.
- LAMAS ALONSO, Carlos: "Eso de las evaluaciones no lo veo claro". Ponencia presentada en el 9º Seminario sobre Audiencia de Televisión (62º Seminario AEDEMO). Lisboa. Febrero 1993.
- LAMAS ALONSO, Carlos: "Eso de las evaluaciones no lo veo claro". *Investigación y Marketing*. Asociación Española de Estudios de Mercado, Marketing y Opinión (AEDEMO). Nº 42. Julio 1993. Págs. 58-62.
- LECKENBY, John D. and BOYD, Marsha M. "How Media Directors View Reach/Frequency Model Evaluation Standards". *Journal of Advertising Research*. Vol. 24. Nº 5. Octubre/Noviembre 1984. Págs. 43-52.
- LECKENBY, John D. y KIM, Heejin. "How Media Directors View Reach/Frequency Estimation: Now and Decade Ago". *Journal of Advertising Research*. Vol. 34. Nº 5. Septiembre-Octubre de 1994. Págs. 9-21.

- LEE, A. M. y BURKART, A. J.: "Some Optimization Problems in Advertising Media Planning". *Operational Research Quarterly*. Vol. 11. Nº 3. Septiembre 1960.
- LEÓN, José Luis. "El Modelo Multinomial Compuesto Dirichlet. Un modelo para calcular la distribución de exposiciones de un plan de medios más exactamente". *Investigación y Marketing*. AEDEMO. Nº 21. Julio de 1986. Págs. 27-30.
- LITTLE, John D.C. and LODISH, Leonard M.: "A Media Planning Calculus". *Operations Research*. Vol. 17. Enero-Febrero 1969. Págs 1-35.
- LITTLE, John D.C. and LODISH, Leonard M. "A Media Selection Model and Its Optimization by Dynamic Programming". *Industrial Management Rev.* Vol. 8. 1966. Págs. 15-24.
- LODISH, Leonard M. "Considering Competition in Media Planning". *Management Science*. Vol. 17. Nº 6. Febrero 1971. Págs. B-293-B-306.
- LODISH, Leonard M. "Empirical Studies on Individual Response to Exposure Patterns". *Journal of Marketing Research*. Vol. 8. Mayo de 1971. Págs. 212-8.
- LOZANO CHOZA, Domingo. "El Estudio General de Medios (II)". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios-Televisión. Gandía. Febrero 1985.
- LOZANO CHOZA, Domingo. "Probabilidad de la audiencia de los medios de comunicación". *Investigación y Marketing*. Nº 19. Noviembre 1985. Págs. 23-6.
- MARC, Marcel: "Combining Simulation and Panel Data to obtain Reach and Frequency". *Journal of Advertising Research*. Vol 8. Nº 2. Junio de 1968. Págs. 11-6.
- MARC, Marcel. "Étude de la Couverture Réalisée par Plusieurs Revues Techniques sur une Clientèle Spécialisée". *Institut de Recherches et d'Études Publicitaires*, estudio nº6. Noviembre 1962.
- MARC, Marcel. "Net Audiences of French Business Papers: Agostini's Formula Applied to Special Markets". *Journal of Advertising Research*. Marzo 1963. Vol. 3. Nº 1. Págs. 26-29.
- MARTÍN ARMARIO, Enrique. "La gestión publicitaria". Ed. Pirámide. Madrid, 1980.
- MARTÍN-GUZMÁN, M. PILAR y MARTÍN PLIEGO, F. JAVIER. "Curso Básico de Estadística Económica". Editorial AC. 3ª edición. 1989.

- MARTÍNEZ RAMOS, EMILIO. "El uso de los medios de comunicación en marketing y publicidad". Ed. Akal/Comunicación. Madrid, 1992.
- MARTÍNEZ RAMOS, Emilio. "Problemas en la estimación de la audiencia de las revistas y diarios en una encuesta tradicional". Ponencia presentada en el 2º Seminario sobre Medios Impresos (54º Seminario de AEDEMO). Madrid. Diciembre 1990. Resumen de ponencias. Págs. 7-33.
- MAULDON, J. G. "A Generalization of the Beta Distribution". *Annals of Mathematical Statistics*. Vol. 30. Págs. 509-20. 1959.
- MAZO DEL CASTILLO, Juan Manuel. "Las decisiones sobre medios en comunicación por objetivos". Tesis doctoral, dirigida por Dr. D. Justo Villafañe. Dpto. de Comunicación Audiovisual y Publicidad I. Universidad Complutense de Madrid. Madrid, 1991.
- MÉNDEZ RODRÍGUEZ, JUAN LUIS. "La medición de la audiencia de radio y sus alternativas". Audiencia y Programación. RTVV. Valencia. 1993. Págs. 99-112.
- MÉNDEZ RODRÍGUEZ, JUAN LUIS. "Método del recuerdo. Estudio General de Medios". Ponencia presentada en el 1º Seminario de Audiencia sobre Radio (49º Seminario de AEDEMO). Madrid. Noviembre 1989.
- METHERINGHAM, Richard A. "Measuring the Net Cumulative Coverage of a Print Campaign". *Journal of Advertising Research*. Diciembre 1964. Vol. 4. Págs. 23-28.
- MORGENSZTERN, Armand. "Frequentation de la Presse, Generalisation, Effet Publicitaire". *IREP*. Etude N° 19. París. Abril 1970.
- MOSIMANN, J. E. "On the Compound Multinomial Distribution, the Multivariate β - Distribution, and Correlations among Proportions". *Biometrika*. 1962. Vol. 49. Págs. 65-82.
- NIELSEN, A. C., COMPANY. "Nielsen Television Index". 1959-64.
- NIELSEN/REPRESS. "Anuario de la Publicidad: 1994". Ed. A. C. Nielsen Company, S.A. Madrid.
- PELLICER, José. "Panorámica actual". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985. Págs. 248-9.
- PISTONI DURÁN, Luis. "El tratamiento de la información para la planificación de medios". Ponencia recogida en: Audiencia y Programación. Editado por el Ente Público RTVV. Valencia. 1993. Págs. 85-97.

- PISTONI DURÁN, LUIS. "Programas actuales". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985.
- PULIDO SAN ROMÁN, Antonio. "Evolución y Revolución del Marketing". Asociación para el progreso de la dirección. Madrid. 1974.
- RAMOND, Charles K. "Agostini's Constant". *Journal of Advertising Research*. Marzo 1961. Vol 1. Nº 3.
- RAY, M.L. and SAWYER, A.G. "Behavioral Measurement for Marketing Models: Estimating the Effects of Advertising Repetition for Media Planning". *Managemet Science*. Vol. 18. Nº 4. Diciembre 1971. Págs. P-73-P89.
- REAL ACADEMIA DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES. Vocabulario Científico y Técnico. Espasa Calpe. Madrid 1990.
- RICE, Marshall D. y LECKENBY, John D. "Predicting Within-Vehicle Television Duplication". Proceedings of the American Academy of Advertising. 1984.
- RUST, Roland T. y KLOMPMAKER, Jay E. "Improving the Estimation Procedure for the Beta Binomial TV Exposure Model". *Journal of Marketing Research*. Vol. 18. Noviembre de 1981. Págs. 442-8.
- RUST, Roland T.; ZIMMER, Mary R.; y LEONE, Robert P. "Estimating the Duplicated Audience of Media Vehicles in National Advertising Schedules". *Journal of Advertising*. Vol 15. Nº 3. 1986. Págs. 30-7.
- SANZ DE LA TAJADA, Luis Ángel y RUBIRA BLÁZQUEZ, Francisco. "Tecnología de los medios publicitarios".
- SANZ DE LA TAJADA, Luis Ángel. "La publicidad: una técnica al servicio de la acción comercial". Ediciones ESIC. Madrid, 1975. Pág. 204.
- SCHREIBER, Robert J. "Instability in Media Exposure Habits". *Journal of Advertising Research*. Vol. 14. Nº 2. Abril 1974. Págs. 13-7.
- SCHREIBER, Robert J.: "Probability Assignments for the Simulation of Media Reach and Frequency". *Journal of Advertising Research*. Junio 1968. Vol. 8. Nº 2. Págs. 3-8.
- SCHREIBER, Robert J.: "The Metheringham Method for Media Mix: An Evaluation". *Journal of Advertising Research*. Vol. 9. Nº 2. Junio 1969. Págs. 54-6.
- SIMMONS MEDIA STUDIES. "Selective Markets and the Media Reaching Them". New York. 1978.

SIMMONS, W. R. AND ASSOCIATES RESEARCH, INC. "Mass Markets and the Media Reaching Them, 1964". New York: W. R. Simmons and Associates Research, Inc, 1964.

SIMMONS, W. R. and ASSOCIATES RESEARCH, INC. "Mass Markets and the Media Reaching Them- 1966 Magazine Reach and Frequency Report". New York. W. R. Simmons. 1966.

SIMMONS, W. R. AND ASSOCIATES RESEARCH, INC. "Selective Markets and the Media Reaching Them, 1964". New York: W. R. Simmons and Associates Research, Inc, 1964.

SIMMONS, W.R. & ASSOCIATES RESEARCH, INC. "1969 Magazine Audience Report". Vol. 1. New York. W. R. Simmons & Associates Research, Inc.

SISSORS, Jack Z. and BUMBA, Lincoln: "Advertising Media Planning". NTC Business Books. NTC Publising. 3ª Edition. 1990.

SOFRES (Société Française d'Enquêtes par Sondage): "Etude des 3 Supports". Paris. 1965.

STARCH, Daniel and STAFF, Inc. *Starch Magazine Report*. New York, 1967.

STEYN, H. S: "On the univariate series $F(t) = F(a, b_1, b_2, \dots, b_k, c, t, t^2, \dots, t^k)$ and its applications to probability theory". *Nederland Akademia Wetensch. Proceeding Series A*, 58. Págs. 588-95.

SURMANEK, Jim. "Una guía fácil y rápida para la Planificación de Medios". Ed. Eresma. Madrid, 1982.

SUSO, Javier. "La función tiempo en la planificación". Ponencia presentada en el 32º Seminario de AEDEMO: Explotación de los Estudios de Medios: Televisión. Gandía. Febrero 1985.

TOM-MICRO: Generación de Resultados. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A. 1995.

TOM-MICRO: Manual de Referencia. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A. 1994.

TOM: Programas para la Planificación de Medios. ODEC, Centro de Cálculo y Aplicaciones Informáticas, S.A.

ZAMBRANO, LOURDES. "La probabilización en el Estudio General de Medios". Ponencia presentada en el 2º Seminario sobre Medios Impresos (54º Seminario de AEDEMO). Madrid. Diciembre 1990. Resumen de ponencias. Págs. 35-57.

Advertencias sobre el diskette anejo

La tesis se acompaña de un diskette (3.5") que sustituye a los habituales anexos: con ello se reduce el papel impreso y se facilitan futuros análisis sobre soporte informático. Los ficheros están en formato WK1, propio de los programas LOTUS, pero convertible a las principales hojas de cálculo. El diskette contiene estos tres ficheros:

- **ALCANCE.WK1**: la hoja se organiza en filas que corresponden a las 402 campañas *multi-soporte* y *uni-periodo* (**II Parte, §A.1.4**); las columnas recogen las características de las campañas, el alcance bruto, la suma de duplicaciones, el alcance neto (según el TOM2) y los alcances netos estimados por los modelos testados en la **II Parte**.

- **DISTRIB.WK1**: horizontalmente se sitúan las primeras 375 campañas (**II Parte, §B.1.4** y **B.9.4**); verticalmente se distribuyen las frecuencias de cada contacto, según estimaban los modelos TOM2, Binomial y Metherringham.

- **TESPANEL.WK1**: incluye los resultados del experimento de la **I Parte, §D**. En filas se organizan las 400 campañas, y en columnas las características de las campañas y cuatro distribuciones de contactos (las observadas en el panel y las estimadas por los modelos TOM, Metherringham y MARKSEL).

Las audiencias y duplicaciones de las revistas, la composición de las campañas, el panel, los ficheros de cálculo, etc, no se incluyen aquí. Harían falta bastante diskettes para contener tal volumen de información. Tengo estos datos disponibles para quien quiera usarlos, siempre con el carácter confidencial exigido por sus propietarios.

Reunido el Tribunal que suscribe en el día
de la fecha, acordó calificar la presente Tesis
Doctoral con la censura de ~~NO ADMITIDA~~ POR UNANIMIDAD

Madrid, 24-11-1995.

EL PRESIDENTE

J. Alonso

EL SECRETARIO

Agel Ferrer
S. A.

VOCAL

~~Agel Ferrer~~

Agel Ferrer

Agel Ferrer